COGNITIVE MECHANISMS OF CATEGORISATION

A Thesis

submitted in partial fulfilment

of the requirements for the Degree

of

Doctor of Philosophy in Psychology

by

Mark Chignell

University of Canterbury

1980

TABLE OF CONTENTS

CHAPTER PAGE

CHAPTER PAGE

ABSTRACT

Purposive behaviour requires the learning of appropriate knowledge about the environment. Cognitive theory and techniques used in mathematical psychology are combined in an exploratory study of aspects of *a* theory of learning. A single quantifiable set of stimuli is used as a preliminary test of the theory and data analysis techniques. A set of 34 features are derived for the stimuli, including scales of complexity and preference. Pairwise similarity judgements are used to assess which of these features are most salient to the perceiver. A delayed similarities task is outlined as a special case of a generalised similarities paradigm. Results obtained in a delayed similarities experiment indicate a shift in the salience of features as compared with corresponding feature salience in a standard similarity judgement experiment. This shift in feature salience is also found when selective attention instructions are given before the similarity judgements are made.

Sorting tasks are used to indicate how participants organise (categorise) the Walsh stimuli.

The results taken as a whole provide basic information about the way in which the Walsh stimuli are perceived and organised.

The present findings need to be checked for their generality using related visual stimuli. Further research within the type of framework to be presented here may eventually lead to a comprehensive theory of learning with respect to two dimensional black and white stimuli.

CHAPTER 1

INTRODUCTION

The research reported in this thesis used a single set of geometric stimuli which were presented in a number of different experimental paradigms. The overall aim of this thesis was to develop experimental paradigms and methods of analysis which would be suitable for the quantitative modelling of the perception, and learning, of visual stimuli.

The organisation of the nine chapters will now be described. Chapter one outlines a cognitive framework within which the experimental results reported in later chapters can be interpreted. Chapter two develops a set of features which can be used to quantify the Walsh stimuli. (The Walsh stimuli are a set of black and white checkerboard patterns which were used as stimuli in all the experiments to be reported in this thesis). Chapter three extends the feature set derived in Chapter two to include the notion of preference, and indicates the existence of subgroups of individuals which differ with respect to their patterns of preference for the various Walsh stimuli. Chapter four investigates the effect of attention on perceptual and cognitive processing. Chapter five studies the way in which the Walsh stimuli are perceived using a standard similarity judgment paradigm. Chapter six introduces the delayed similarities paradigm and relates it to other modifications of the basic similarity judgment paradigm. Chapter seven looks at some specific issues in more detail, including the relationship between similarity judgments and reaction times. Chapter eight uses sorting tasks to study the way in which the Walsh stimuli are grouped together. Conclusions drawn from the first eight chapters, along with the prospects for future research, are given in Chapter nine.

The remainder of this chapter will consist of a general metatheoretical frame work within which the results to be reported in succeeding chapters may be interpreted, followed by a brief introduction to the stimuli, and other resources, that were used.

Knowledge and Learning

Knowledge may be regarded as the goal of learning, and the need for humans to gain appropriate knowledge of their environment makes learning perhaps the most fundamental psychological process. In view of this it is not surprising that the various schools in psychology may still be characterised largely in terms of the respective theories of learning that they espouse. The one unifying concept is that learning is evidenced by changes in behaviour.

> "Learning is the process by which an activity originates or is changed through reacting to an encountered situation, provided that the characteristics of the change in activity cannot be explained on the basis of native response tendencies, maturation, or temporary states of the organism (e.g., fatigue, drugs, etc)."
>
> —Hilgard and Bower (1966, p.2).

Thus, in attempting to understand the learning process it is necessary to deduce changes in the knowledge that a person has by observing changes in that person's behaviour. Such understanding requires:

(i) A theory relating knowledge to its behavioural consequences.

(ii) Experimental paradigms which are able to test and refine theories of knowledge and its acquisition.

The present chapter willl use a number of concepts from the field of cognitive psychology in developing a theory of learning which will provide the necessary framework for interpreting the results of later experimentation. One such concept consists of a knowledge structure which is formed during the acquisition stage of learning and utilised in goal-oriented problem solving. The notion of knowledge structure, however, presupposes some form of cognitive representation.

Palmer (1978) has characterised representation in general as follows:

"The nature of representation is that there exists a correspondence (mapping) from objects in the represented world to objects in the representing world such that at least some relations in the represented world are structurally preserved in the representing world.

In other words, if a represented relation R, holds for ordered pairs of represented objects, $< x , y >$ then the representational mapping requires that a corresponding relation, R', holds for each corresponding pair of representing objects, $< x', y' >$

— Palmer (1978, pp 266 – 267).

Palmer also states (1978, p.300) that one must first consider the functional information content of an object as defined by the processes that use that information before determining the representational nature of the object. A recent study by Nickerson and Adams (1979) provides a graphic illustration of this point. They tested the long-term memory for a common object of their participants by having them recall a United States one-cent piece or recognise that coin from a number of facsimiles of the one-cent piece. Their participants showed poor recognition and recall performance across all five experiments conducted. For instance, participants were unable to specify what was wrong with erroneous facsimiles of the one-cent piece, and two of those facsimiles were accepted as correct representations of the coin as frequently as was the correct drawing. Nickerson and Adams interpreted their results in terms of the functional information required concerning the coin in everyday life:

"Why are our memory representations for so familiar an object not more complete and precise?

One plausible explanation is that there is no need for them to be any better. Perhaps what we mean when we say that we know what a penny (a U.S. one-cent piece) looks like is that we can distinguish a penny from other things from which we normally have to distinguish it, for example, from other coins . . . What is interesting about this explanation is that it suggests that many of the numerous things we all can 'recognise,' we may recognise on the basis of memory representations that are as incomplete and imprecise as our representations of pennies appear to be."

— Nickerson and Adams,(1979, p. 304).

This functional aspect of representation can be seen as a direct result of the individual's need to gain *appropriate* knowledge of his/her environment (at the beginning of this section).

Stored representations can be used to incorporate past experiences into present perception, whether these representations act as a constructed anticipation of a certain type of information (Neisser, 1976; p.20), or whether they consist in knowing how to utilise cues with reference to a system of categories (Bruner, 1973; p.7). The terms categories and categorisation will be used in this thesis to refer to stored representations and their acquisition. The use of this terminology does not imply one type of theory of cognitive representation (Rosch, 1978; Bruner, 1957), in favour of another type (Neisser, 1976; Gibson, 1966). In many cases it is difficult to see how one theorist's concept of category differs from another theorist's concept of schema, let alone design an experimental task which could falsify one type of theory but not the other. It is thus not surprising that a theorist may even use the two concepts interchangeably:-

> "Encoding and storing faces according to a schema simply means that *we need not notice all of the differences among faces in order to tell them apart."*
>
> – Hochberg (1978, p.218).

> "Some of the cues by which we categorise a person are *extrinsic:* e.g., uniforms that reveal official roles and status, . . . But many categories are indicated by the face alone . . . "
>
> – Hochberg (1978, p.219).

The first quotation implies that we categorise different faces as belonging to different people. The second quotation implies that we categorise people into different types. Thus the difference in the categorising process appears to reside in the amount of abstraction involved. Alternatively, the first quotation implies that the schematic control of encoding involves a summation of cue validities (Reed, 1973; p.167) while the second implies that such cues may be culturally determined. The examples given above highlight the fact that differences in the use of the concepts "schema" and "category" are essentially one of emphasis. "Schema" tends to refer to a process (such as cue utilisation) while "category" generally refers to the stored structure which can be used to determine how the process acts (e.g. utilising and weighting cues according to their ability to discriminate between the available categories).

Previous notions of category and schema may thus be incorporated as complementary components in a general theory of knowledge. A complete theory of knowledge would need to consider how knowledge is acquired, stored, and utilised. The somewhat less ambitious aim of this thesis is to provide some of the theoretical and experimental tools needed to study the acquisition, and storage, of knowledge about simple visual stimuli.

Knowledge Acquisition

Knowledge acquisition may usually (though not always) be regarded as an increase in the adaptive fit between the person and his/her environment, where the type of fit is determined by the constraints operating in the environment as well as the purposive orientation of the person. A basic property of such adaptive fitting is that it is a dynamic process occurring through time. Gregson, (1978, 1980) has recently outlined and used techniques for fitting dynamic models to data *within* an experimental session which allow the precise modelling of a specific process. Two processes may well be involved in learning, which I shall call *macroscopic* and *microscopic* learning. Macroscopic learning involves global changes in the stored representation and its effect can generally be estimated by comparing results *between* experiments. In contrast, microscopic learning will involve changes to representational structure. (This important distinction between stored representations and the representational structure generated in a particular experiment is explained in the next section) and may be detected *within* an experiment using methods such as those provided by Gregson.

Under the process of macroscopic learning knowledge is acquired incrementally through successive (and in most cases, increasingly precise) approximations of reality. This incremental approach is necessitated by the physical, intellectual and emotional limitations of people in general (Stotland and Canon, 1972). Recognition of similarities between sets of objects, and the formation of groupings on the basis of similarity in function and appearance, is likely to be part of the incremental process of knowledge acquisition. Such preliminary organisation makes the most of whatever approximate order can be established at a given time and ignores, at least temporarily, complexities which are currently too subtle to be systematised.

This view of similarity as an abstractive process has been previously stated by Gregson:

> "In encoding items into long-term memory . . . items have to be represented and located in such a way that those which are related, by resemblance of meaning, are linked, and even encoded in a hierarchical manner . . . The general idea is that the memory encodes similar items close together . . . so that they may be retrieved by an economical search strategy, and errors of recall will substitute pair-wise similar items."
>
> — Gregson (1975, p.207).

The emphasis in this thesis is on pre-asymptotic stored representations of a stimulus set which will change incrementally over time during the process of learning.

Stored Representations and Representational Structure

A distinction may be made between the representations of stimuli which remain approximately constant over different experimental tasks (once a learning asymptote has been reached) and those representations which change as the experimental task changes. As mentioned previously, stored representations change as knowledge is acquired incrementally over time. Thus the stored representations of a participant during an experiment will depend more on previously accumulated knowledge of the stimuli (providing that his/her familiarity with the stimuli exceeds some minimal level) than on the immediate context provided by the experiment. The notion of representational structure is introduced as a form of representation which is sensitive to changes in the experimental conditions and stimulus context. This distinction between stored representations and representational structure may appear to correspond to the distinction between primary and secondary memory (Waugh and Norman, 1965) but there are reasons for using the terminology presented here which will be outlined in chapter four. Given that the participant has semi-permanent stored representations, and a representational structure built up during the course of the experiment, some account still needs to be taken of present, and very recently past, perception. For the purposes of this discussion stimulus input and task demands may be viewed as acting on the stored representations to produce activated memory schemata (Norman and Bobrow, 1976). The activated memory schemata will change from trial to trial, but some process (possibly averaging) will produce a representational structure out of the succession of activated memory schemata. The notion of representational structure outlined here is similar to Helson's (1964) 'Adaptation-Level Theory' except that the adaptation here is not in terms of the magnitudes of stimulus properties but rather in terms of the inter-relations of stimuli. In terms of their predictions of judgmental elicited in psychological tasks, there may not be much difference between the concepts of adaptation level and representational structure. In any case, the term representational structure will be used here as it emphasises the cognitive rather than the psychophysical aspect of experimental tasks.

Theories and Models

There appears to be a useful distinction between theories and models which has been summarised by Palmer (1978):

"A theory of something is essentially a description of it at some level of analysis. It expresses the structural laws that hold in the object of study at a level of abstraction appropriate for the goals and methods of the scientific enterprise for which it is constructed. A theory, then does not include aspects that are more concrete than can be verified by empirical observations of the sort indigenous to the science. A model is a concrete embodiment of a theory. Its relationship to its theory is that it satisfies the assumptions of the theory. Because there are many ways in which a given theory may be satisfied, there are many models that are consistent with it."

— Palmer (1978, p.275).

The theory of learning presently being outlined is at a high level of generality because of the high level of generality implied by the type of observations relevant to human learning which psychologists are currently able to make. One method of checking the plausibility of theory is to specify particular models (based on the theory) which can then be assessed for the predictions they make under various simulated conditions. Model simulation of the present theory will not be carried out in this thesis as the main aim here is to increase the concreteness of aspects of cognitive theory that can be verified by empirical observations (cf the quote above) by careful modification of presently available experimental paradigms. It is recognised, however, that the theory presented here should ultimately be tested for its plausibility and internal consistency using simulation techniques.

Knowledge Representation During Acquisition

Given an incremental process of knowledge acquisition the apparent representational structure of the knowledge at any time increment should depend on the experimental paradigm used in exploring that structure. The idea that type of processing is related to the perceptual operations required by the experimental task has previously appeared within the framework of "levels of processing."

"If the memory trace is viewed as the by-product of perceptual analysis, an important goal of future research will be to specify the memorial consequences of various types of perceptual operations."

— Craik and Lockhart (1972, p.681).

Peterson (1977) has reviewed some of the studies which provide support for the levels of processing hypothesis. Baddeley (1978) on the other hand, has suggested that the levels of processing framework be abandoned in favour of an approach that explores specific components of the memory system. In the present thesis, the idea of different processing levels is compatible with the theory being outlined. However, the view here accords with the notion of levels of retrieval processing, whereas it is encoding processing which is the focus of current approaches to levels of processing (Craik, 1979). As a consequence, most of the arguments in the current debate over the usefulness of levels of processing as a metatheoretical framework are not relevant to the present analysis and will not be considered further. The view followed here is that task demand characteristics determine the type and complexity of processing carried out on the stored representation of a stimulus (or stimuli) at a given time increment. The resulting representational structure at that time increment will be a concatenation of the present stimulus input and the stored representation. This view enables us to operationalise the concept of levels of processing in retrieval in terms of experimental tasks which require different retrieval strategies, an approach which has been suggested previously:-

". . . We can explore the correspondence between the processing demands of a task and the strategies for its optional performance.

A body of knowledge about this correspondence between tasks and optional strategies for humans, could, in turn, serve as the groundwork for two related projects: first, it could provide the basis for classifying tasks in terms of the types of information processing used by experienced subjects while performing them . . . and second, it could enable researchers to develop the ability to decompose a task into a number of basic information-processing requirements and, from that analysis to predict the strategies that would allow best task performance."

— Gilmartin, Newell and Simon (1976, p.16).

A corollary of this view is that the apparent knowledge representation during acquisition will depend on the experimental tasks used, as well as the amount, and type, of previous experience of the stimuli.

The Effect of Task Demand Characteristics

The effect that task demand characteristics have on the participant's responses will be discussed in Chapter four where it is suggested that task demand effects are mediated by changes in attentional strategy. For now it may be noted that task demands will change the relative usefulness (or salience) of information from the stored representations and/or the representational structure. The Nickerson and Adams (1979) study mentioned earlier may be regarded as involving an extreme form of task demand where the information relevant to the coin was changed from an assessment of its relative monetary value to the (previously non-functional) enumeration of the features impressed on the coin. The next section will consider how experimental responses are produced.

Response Derivation

One view of responses is that they are selected from a finite set of existing alternatives, presumably according to some implicit pattern recognition process where a given stimulus pattern is mapped into a response category. Another view (the one taken here) is that the apparent selection of response categories is as an artifact of the way the participant is required to communicate his/her responses to the experimenter (i.e. making a numerical rating, or in general, choosing one of a set of possible response categories which have been previously defined by the experimenter).

Instead of this view of responding as a selection process, it is assumed here that the requirements of the task make the participant impose constraints on the form of the representational structure of the stimulus set. It is of course possible, particularly with overlearned stimuli, that the participant bases his judgment directly on the stored representation. The effect of task demands on the stored representations vis a vis the representational structure is a matter for empirical investigation and will be considered again in the light of the experiments to be reported in later chapters. For the present discussion it will be assumed that the overt responses made by a participant during an experiment are derived by operations upon the representational structure in order to make the responses (e.g. ratings) which are appropriate within the context of the experiment.

One possible mechanism for response derivation in a similarities experiment involves computing the distances between a representation of the stimuli in perceptual space (which is presumably approximated by the multidimensional scaling solution derived from the similarity responses) and converting these into similarities. This and other models of response derivation (set-theoretic models of similarity judgment in particular) will be considered in Chapter five.

Related Psychological Theories

Many aspects of the theory presented so far are, of course, closely related to theories already available in psychological literature. However no single current theory appeared to be wide enough in scope to provide a framework for the investigation of human learning. Consequently the present theory is being developed to give at least a minimal guideline for investigation in the area.

Two of the theories which are closely related to the present conceptualisation are Kahneman's (1973) schematic model (this is in fact a *theory* according to Palmer's (1978) criteria) for attention and perception, which is illustrated in figure 1.1 and Norman and Bobrow's (1976) 'memory schematic' view of the human information processing system.

The main difference between Kahneman's theory and the one used in this thesis is that the present theory is cognitive in orientation whereas Kahneman's theory is oriented more towards perception. In the type of experimental paradigm to be used in this thesis (where all the stimulus information may be scanned within the time available) the processes between sensory registration and the activation of recognition units (or, in the terms of the present theory, the activation of memory schemata within the stored representations) is assumed to occur automatically (Deutsch and Deutsch, 1963; Keele, 1973; Shiffrin and Schneider, 1977). Thus the effect of task demands will be assumed to occur *after* the activation of memory schemata. The other result of Kahneman's perceptual orientation is his use of the notion of response selection rather than the notion of response derivation used here.

As well as assuming that a memory schema is activated by sensory input ('bottom-up processing') Norman and Bobrow (1976) also incorporated 'top-down processing':

"the schema representing an object within the memory system must contain information about the relationships among the parts of the object, and whenever some corresponding relationships or identifiable subparts are observed in a perceptual input, that schema must be activated as a potential organisation for the input."

— Norman and Bobrow (1976, p.120)

Within the present theory top-down processing can be incorporated as the effect of representational structure and/or stored representations on processing. Such top-down processing may have an important effect on the delayed similarity judgements used in the experiments outlined in Chapter six.

Figure 1.2 gives a schematic representation of the type of theory considered above, outlined in a similar form to that of figure 1.1. The theory and its representation will be revised in Chapter four.

The remainder of this chapter will introduce the stimuli used in the experiments to be described below, as well as giving a few technical details. Chapter two will derive a set of features which can be used to quantify the stimuli, and these will be used to interpret later experimental results.

Stimuli

A single set of 64 visual stimuli was used in the studies to be reported here. The stimuli are given in figure A.1 of Appendix A. The derivation of these stimuli, and some of their mathematical properties, are given in Appendices A and B. The Walsh stimuli were chosen because they were both quantifiable and novel (the author avoided the use of electrical engineers as participants in his experiments). In addition, they are an intrinsically interesting set of quantifiable stimuli as their orthogonality makes some of the models previously used in quantitative psychology inappropriate (See Appendix A).as models of similarity judgments between Walsh stimulus pairs.

Apparatus

All the pairwise similarity experiments reported below utilised the following equipment:- Two Kodak carousel slide projectors controlled by a PDP11/10 (Bell et al, 1978) mini-computer. Similarity rating responses were made by pressing one of up to 10 response buttons mounted in a board which was placed in front of the subject (participant). Each trial of a particular experiment consisted of the presentation of a pair or slides (Walsh stimuli) followed by the storage of the rating response and the reaction time for that response.

Participants

The participants in the series of experiments reported in this thesis were all psychology students at the University of Canterbury, none of whom had had any previous experience with the Walsh stimuli.

Literature

A large portion of the relevant research literature was continuously surveyed up until

November 1979. In general, studies will only be mentioned if they are directly relevant to the line of argument being pursued.

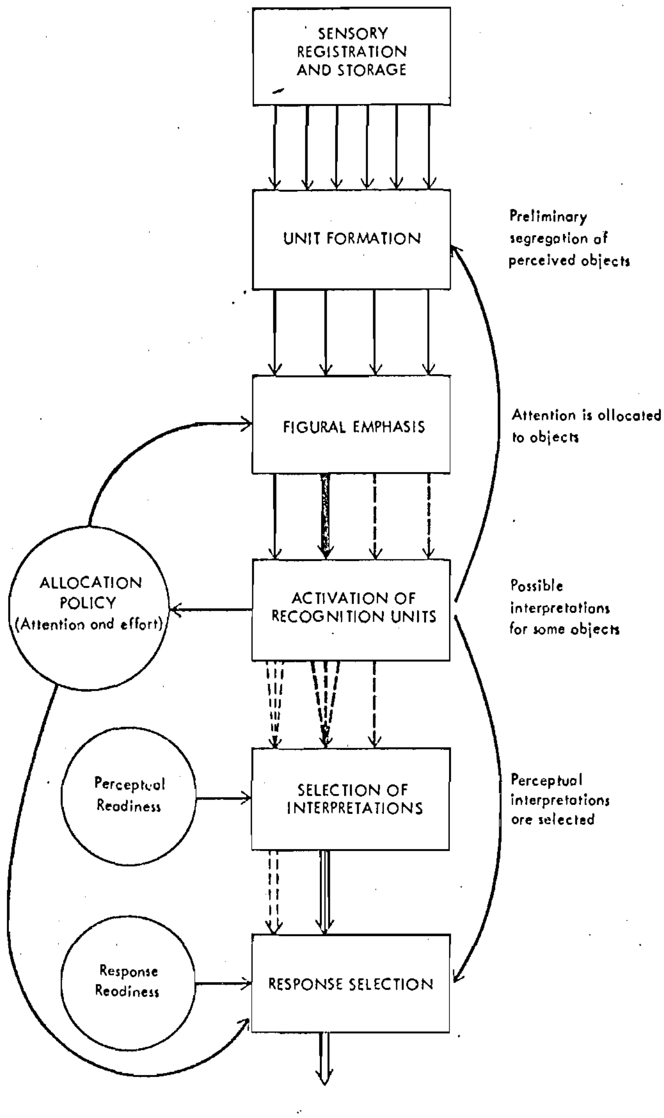


Figure 1.1 A schematic model for attention and perception (After Kahneman, 1973, Figure 5.1).

Computer Programs

The computing for this thesis was performed on either the Burroughs B6718 of the Computer Centre, University of Canterbury, or the PDP11/10 of the Psychology Department, University of Canterbury, using a variety of local and published programs. The published programs used will be referenced in the text where they are mentioned.

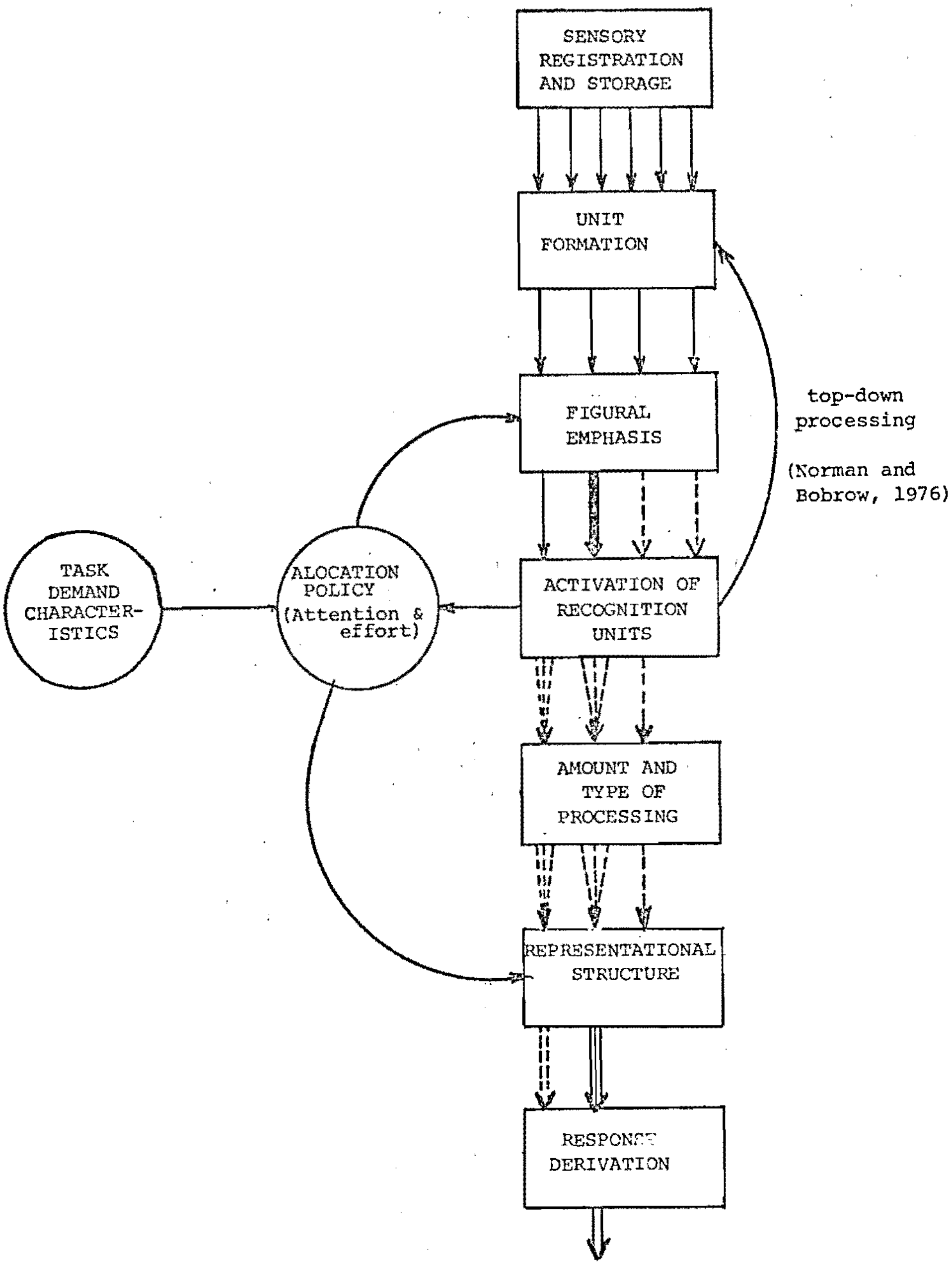


Figure 1.2 A schematic representation of task demands effects on attention and perception.

CHAPTER II

In order to quantify perceptual learning of the Walsh stimuli it is necessary to determine what the quantifiable aspects (features) of those stimuli (which are relevant to their perception) are. Zusne (1970, Chapter Five) has summarised the features which have been previously used in the study of visual form perception. Some of the features of the Walsh stimuli derived later in this chapter will be based on the types of feature outlined by Zusne, but many of the features derived previously are not appropriate for the Walsh stimuli as they were often developed with respect to random shapes (Attneave and Arnoult, 1956) which are generally outline figures. Even when stimuli are quantifiable in terms of a set of features there is no guarantee that these features will be perceived as such by the human observer, since there is as yet no completely satisfactory theory of what the subjectively perceived properties of visual stimuli are:

"Human 'pre-processing' or feature selection is a highly complex and apparently nonlinear operation which has not, as yet, been duplicated through mathematical formulation . . . He (a human) evinces a heuristic feature selection based upon his past experience, and the selection process is difficult if not impossible to describe as he himself often cannot explain upon what specific attributes he based his decision."

— Andrews (1972, p.16).

In the absence of a satisfactory theory of human feature selection the present chapter will derive as many features of the Walsh stimuli as possible. The question as to which of these features are relevant to people's judgments of the Walsh stimuli will then be answered by referring to the experimental results described in later chapters. One problem with this approach is that there is no simple way of deciding which of a set of correlated features are in fact the psychologically relevant features. Previous approaches have often assumed that features are hierarchically related within a pattern recognition process (e.g. Neisser, 1967). In this view features are correlated because they are nested and selection of the best feature needed to explain the experimental results may be aided by noting the implications of the hierarchical theory. Henderson (1978) has persuasively criticised the hierarchical approach to pattern recognition. However, whether features are or are not hierarchically organised, we expect that differences in responding at different stages of an experiment should relate in some way to quantifiable aspects of the stimuli used. The ensuing problem of deciding which of the quantifiable aspects (that are related to differences in responding) are psychologically relevant will be considered in later chapters (after the appropriate experimentation has been reported).

Two conceptually separate sets of features may be distinguished. The first set consists of features which are based on the physical properties of the Walsh stimuli. The second set consists of features which are derived from psychological scaling of the Walsh stimuli. Both stimulus sets are likely to be necessary for an adequate account of the perception of the Walsh stimuli:

"There is a necessary mutual interplay between stimulus and organismic properties. An organism cannot engage in pattern recognition, for example, based on a feature analysis unless there are in fact features in the stimulus to be analyzed. On the other hand, there is no need to attempt pure stimulus descriptions in terms that are inappropriate to the processing organism. So the properties of the organism limit the properties of the stimulus to which we pay attention; at the same time, the properties of the stimulus limit what the organism can do with the stimulus."

— Garner (1978, p.101).

1. Features based on physical properties of the Walsh stimuli

A total of 23 features were selected. The first 11 features were derived from a close examination of the quantitative structure of the Walsh stimuli, in view of previous stimulus variables used with black and white geometrical stimuli (Brown and Owen, 1967; Zusne, 1970).

Feature 1 Column Sequency[1] (Appendix A, p.16).

Feature 2 Row Sequency (Appendix A, p.16).

Feature 3 Block Structure (Appendix A, p.13).

Feature 4 The number of black squares contained in the stimulus.

Feature 5 The number of black rectangles contained in the stimulus.

Feature 6 Average grain, defined as:

(2 x perimeter)/(row sequency x column sequency)

where the perimeter is the sum of the perimeters of all the black shapes (rectangles or squares) which make up the stimulus.

Feature 7 Grain variance, defined as:

$$\sum_{S} [(P_s - A.G)^2]/m(s)$$

where S is a particular black shape, m(s) is the total number of black shapes in the stimulus,[2] P_s is the perimeter of a particular black shape and A.G is the average grain (feature 6, above).

The derivation of some of the remaining features may be clearer if the schematic representation given in Appendix B is consulted. Using this system, stimulus number 38 (Walsh 5,6) may be represented as in figure 2.1.

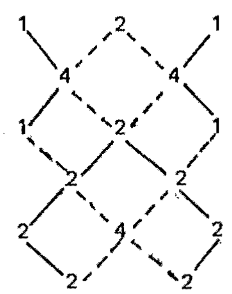


Figure 2.1: Schematic Representation of Wal (5,6)

1. In simple visual terms, column sequency is the number of vertical stripes in a Walsh stimulus, while row sequency is the number of horizontal bands in the stimulus. These bands and stripes are harder to detect as the row and clumn sequencies get larger, however an inspection of Figure A.1 (Appendix A) should provide an impression of what sequency involves.

2. M(s) is the sum of the number of squares (feature four) and the number of rectangles (feature five).

The numbers in figure 2.1 indicate the areas of the shapes while the connecting lines (edges) indicate when two shapes are adjacent. If the connecting line is continuous then the two shapes are similar, whereas a dashed connecting line indicates that the shapes are different (one a square, the other a rectangle).

Feature 8 Heterogeneity of Area. If we consider figure 2.1 as a graph, then each shape is a node with a cardinality (N_s) equal to the area of that shape.

We now define heterogeneity of area (HA) as follows:

$HA = \Sigma [N_s \text{ (left node)} - N_s \text{ (right node)}^2] / m(e)$

where m(e) is the number of edges in the stimulus.

Feature 9 Heterogeneity of Shape. For each edge i, we construct a measure m_i where m_i equals 1 if the connecting edge is continuous (same shaped nodes) and 0 if the edge is broken (different shaped nodes).

Heterogeneity of shape (HS) is then defined as:

$MS = (\sum_e Mi) / m(e)$

Feature 10 Squareness (SQ).

SQ = Number of squares / m(s)

where m(s) is the number of shapes (cf feature 7).

Feature 11 Combined Hetergeneity (CH).

CH = HA x HS

where HA and HS are as given for features eight and nine respectively.

The remaining twelve features have been mentioned previously in the engineering and pattern recognition literature. Haralick et al. (1973) considered the problem of deriving textural features for the purpose of image classification. The fundamental unit of textural analysis is the spatial dependency matrix. Twelve spatial dependency matrices were derived for each of the 63 non-homogeneous Walsh stimuli (stimulus one was redundant with respect to texture). Five features were calculated on each of the 12 matrices and 12 rotationally invariant features were finally selected from these as a set of textural features for the Walsh masks. A full account of the process of selecting and calculating this set of 12 textural features plus a listing of the BASIC programme used to derive them and the feature scores for each of the Walsh stimuli, are given in Appendix B.

1. m(s) is the sum of the number of squares (feature four) and the number of rectangles (feature five)

2. This notation is explained in Appendix B.

2. Features Based on Psychological Properties of the Walsh Stimuli

The preceding section outlined 23 features which represent objective properties of the Walsh masks. Such features, by themselves, are unlikely to give an adequate account of the psychological properties of the Walsh stimuli:

"One major criticism of any attempt to study the perception of patterns in terms of their objective properties is that whenever we specify the basic units of which the pattern is composed we do this arbitrarily and can never know what units the perceiver uses . . . if we have constructed the stimulus we know exactly what its components are. Even if the observer does not perceive the stimulus in terms of this set of components the set he does perceive must be some kind of transformation of this original set."

– Frith (1978, p.232).

One subjective feature that has recently been investigated in the perception of checkerboard patterns is their complexity (Chipman, 1977). Other subjectively perceived features which may influence perception are symmetry and preference. Experiments E1, E2 (below) and E3 (in the next chapter) attempted to scale the Walsh stimuli in terms of these subjective features.

Experiment Extraction 1 (E1)

Nine first-year psychology students were used as participants. The 64 Walsh stimuli were shown singly in random sequence. The presentation of this sequence was repeated three times. On each trial of the total 192 trials, the participant was required to rate the stimulus shown in terms of one of three verbal labels. 'Complexity' was the label used for the first 64 trials, 'Symmetry' was used for the second set of 64 trials, and 'Jaggedness' was used for the remaining trials.

The three labels used in E1 (complexity, symmetry and jaggedness) are the names of three objective measures previously found to be important in the perception of visual form (Zusne, 1970). The use of these loosely constrained verbal labels was designed to encourage the participants to scale the stimuli according to the features which were important to them, not the features which were important to the experimenter. Fuzzy cluster analysis (Bezdek, 1974) was used on the data, in order to identify the underlying features being used. The results were not interpretable in terms of the first-order physical properties of the Walsh stimuli, probably because of wide variation between participants in the use of the 10-point rating scale.

A summary of the complexity results ordered according to sequency, is given in table 2.1.

Row Sequency								
8	4.33	6.22	5.44	4.11	4.88	6.56	5.11	7.67
7	5.22	3.89	4.78	5.78	5.11	3.89	5.33	5.33
6	3.78	5.11	4.78	5.22	5.56	3.11	4.22	4.44
5	4.89	3.89	6.67	4.44	4.44	3.78	4.44	4.89
4	5.11	4.56	4.67	5.33	5.89	4.44	3.78	4.67
3	4.56	4.56	3.78	6.22	5.78	3.89	6.00	4.89
2	4.11	6.22	5.44	5.44	5.11	4.67	4.78	4.33
1	3.00	6.33	5.11	5.33	5.33	5.00	5.56	4.67
Column Sequency	1	2	3	4	5	6	7	8

Table 2.1 Mean Complexity Ratings (E1 data) Ordered by Sequency

It can be seen that Table 2.1 is not, in general, symmetrical about the bottom-left to top-right (negative) diagonal. This indicates that the measure of complexity represented in Table 2.1 varies under rotational transformation. It can also be seen that the checkerboard pattern (Wal (8,8)) has the highest mean complexity rating (7.67) which does not conform with the intuitive notion that complexity should be closely related to memorability.

Figures 2.2 a, b, and c show the number of times each participant used each of the 10 points on the rating scale over the three stages of the experiment. The variation both within and between subjects indicates a certain amount of numerical response bias, which makes it inadvisable to group or compare results.

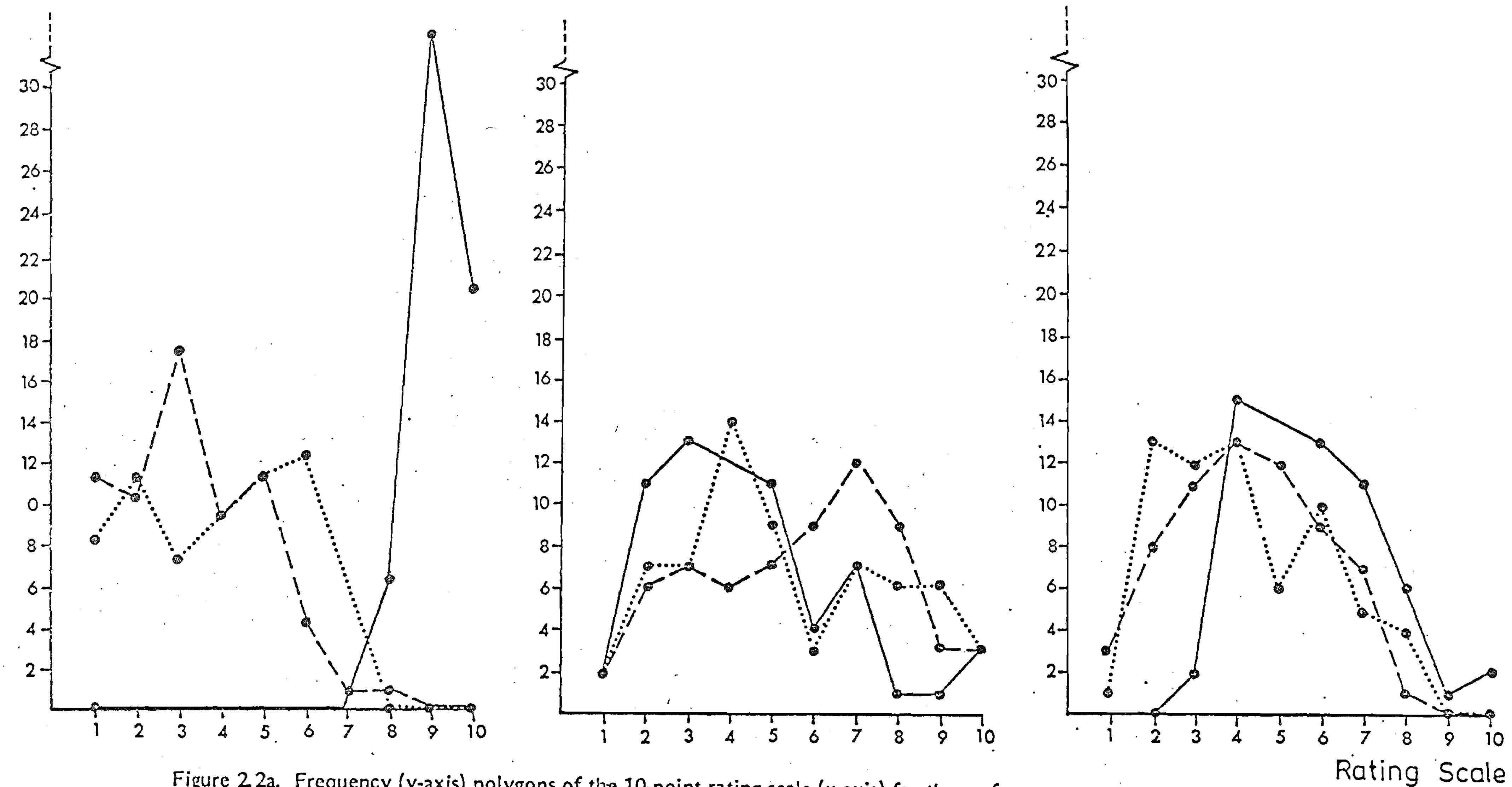


Figure 2.2a. Frequency (y-axis) polygons of the 10-point rating scale (x-axis) for three of the E1 participants.

- complexity.
- jaggedness.
- symmetry.

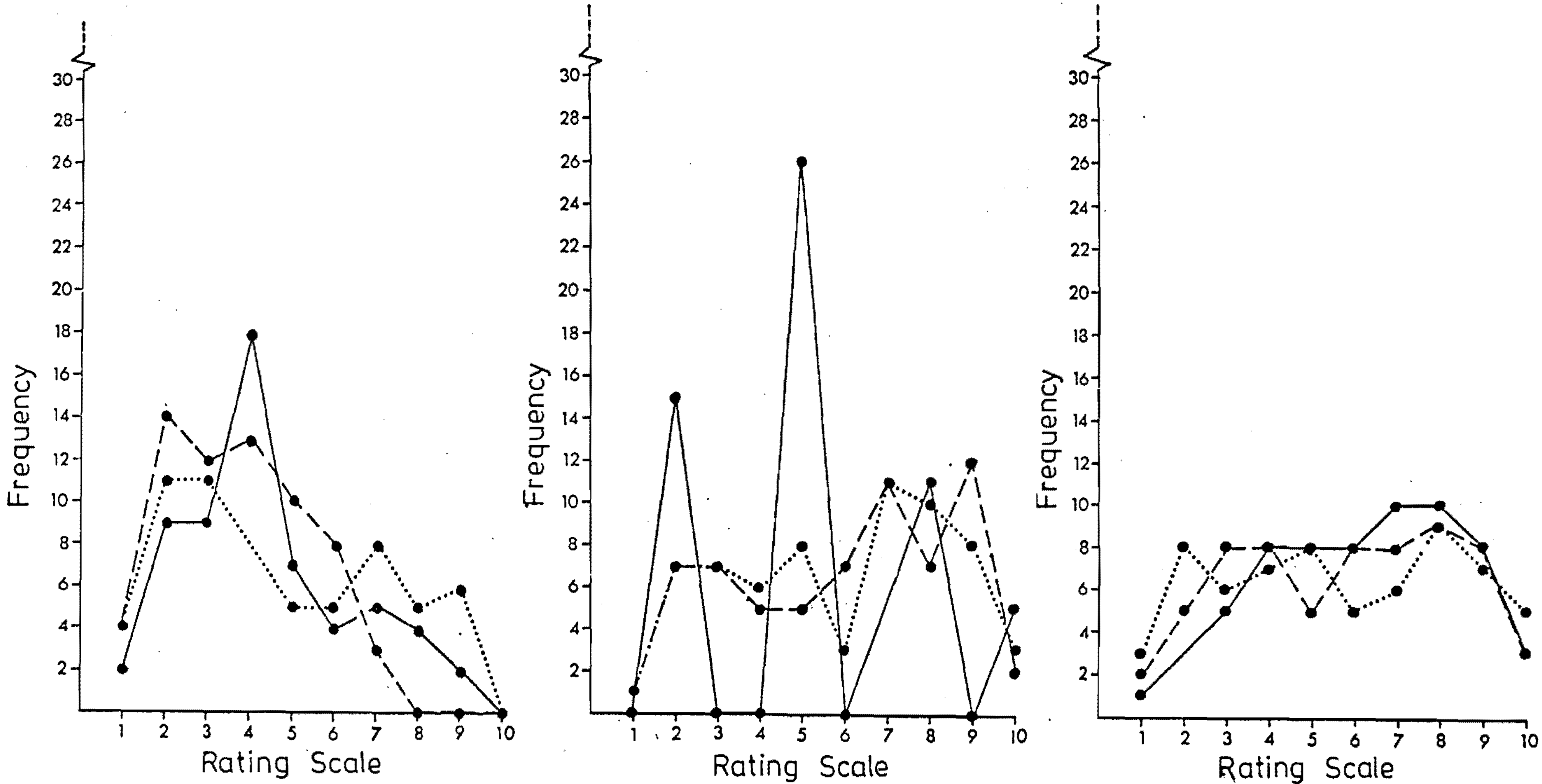


Figure 2.2b. Frequency (y-axis) polygons of the 10-point rating scale (x-axis) for three of the E1 participants.

●·······● – complexity.

●– – –● – jaggedness.

●———● – symmetry.

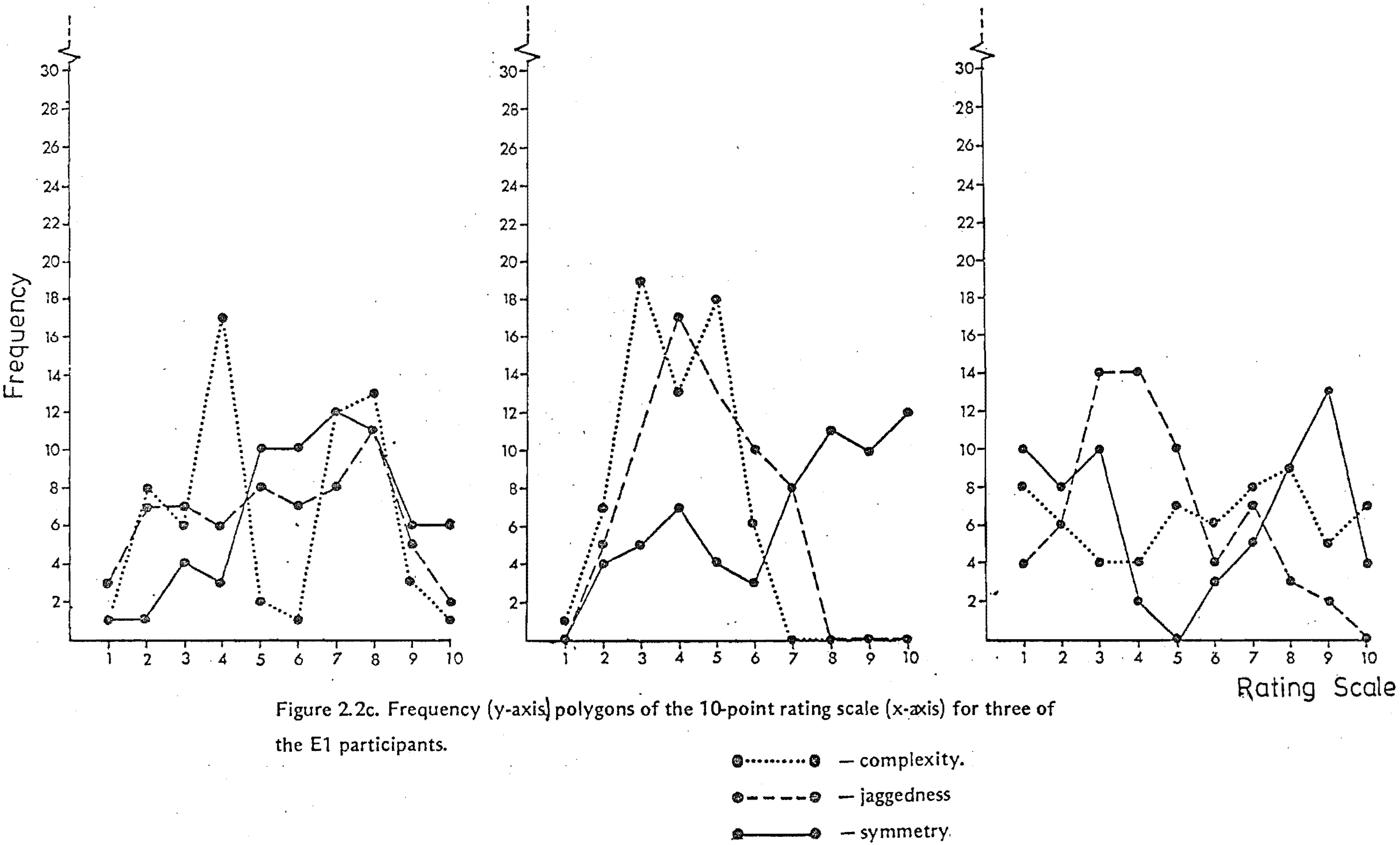


Figure 2.2c. Frequency (y-axis) polygons of the 10-point rating scale (x-axis) for three of the E1 participants.

The properties of the complexity measure in Table 2.1, as well as the problem of individual differences in the use of numerical ratings (Figure 2.2), show the results of E1 to be unsatisfactory. As an alternative I attempted to devise an experimental method which would allow the participant to express his/her judgement of a subjective feature without having to make numerical responses. This method is entitled conceptual ranking. (Methodological issues in its construction and use are discussed in Appendix C.) The experiments outlined below used the method of conceptual ranking to rank the Walsh stimuli in terms of perceived complexity.

Experiment E2

This experiment sought to derive a measure of complexity by using

1. a sorting procedure which did not require numerical responses, and
2. a constrained definition of complexity which would be interpreted in approximately the same way by all of the participants.

Method:

12 participants were each asked to construct a 2-way (8 x 8) conceptual ranking of the Walsh stimuli with respect to how complex each stimulus was. The conceptual ranking procedure is as follows:

The 64 stimuli are randomly placed into the form of an 8 by 8 grid on the top of a large table. The participant then carries out the following procedure.

Step 1.

Rank the first column (working left to right, say) in ascending order (working upwards) according to the value of the concept (complexity).

Then, do the same with each of the seven remaining columns, working from left to right across the columns.

Step 2.

Rank the first (bottom, say) row in ascending order (working to the right) according to the value of the concept. Do the same within each of the seven remaining rows, working up the grid row by row.

Step 3.

If none of the stimuli have been moved since you were last at STEP 3, then Go to STEP 4. Otherwise, return to STEP 1.

Step 4.

Are you satisfied with the configuration?

The result of this procedure is an 8 by 8 grid of the objects which is ranked simultaneously over both rows and columns by the value of the concept.

An important feature of the above technique is that it is designed to converge iteratively on a psychologically stable configuration. It does not, however satisfy mathematical criteria for stability as using the method on a set of 35 ranked numbers placed randomly in a 7 x 5 grid produces configurations which are dependent on the original placement of the numbers in the grid (Appendix C gives further details of this problem).

At present the conceptual ranking method is a useful heuristic method of scaling psychological stimuli, however it is fallible in the sense that a unique set of stimulus rankings on the concept cannot be inferred from the final configuration. It is possible to average out the errors in the conceptual ranking technique by pooling the results across individuals (as shown below). Refinement of the conceptual ranking technique will be necessary, though, before it will be able (at least in terms of the mathematical properties of the ranking process) to accurately identify the scale rankings of a single individual.

The constrained measure of complexity used as the concept in the E1 ranking was defined in the following instruction given to each participant.

"Rank the cards within each row or column according to how long you think you would need to look at each card so as to be able to reproduce that card after it had been removed for a minute."

Deriving the complexity rankings

Once a conceptual ranking has been obtained using the instructions given above, the implied rank of each of the stimuli may be inferred from it. Table 2.2 gives an implied rank for each position in the resulting matrix. The complexity rank for each stimulus was taken to be the implied rank assigned (as in Table 2.2) to the position of that stimulus in the conceptual rank.

8	9	10	11	12	13	14	15
7	8	9	10	11	12	13	14
6	7	8	9	10	12	12	13
5	6	7	8	9	10	11	12
4	5	6	7	8	9	10	11
3	4	5	6	7	8	9	10
2	3	4	5	6	7	8	9
1	2	3	4	5	6	7	8

Table 2.2 Implied Rank by position in an 8 x 8 Conceptual Grid.

- A Measure of Complexity

The results of the conceptual ranking, for a single participant, may be presented as a sequency ordered matrix, where the stimuli are replaced by their assigned ranks. The underlying row and column sequencies define a two-way row x column matrix which is anolagous to an eight by eight design. Analysis of variance (ANOVA) was used to check for heterogeneity in complexity ratings by including the participants as 12 levels of a third (replications) factor. Neither of the replications x row sequency, nor the replications x column sequency, effects was significant ($p < .05$)

As a consequence, the mean ranking (averaged over the 12 participants) of the ratings for the stimuli (Table 2.3) may be used as a general measure of complexity for the Walsh stimuli. This measure of complexity was, of course, derived using a set of instructions which required the stimuli to be rated according to a subjective impression of memorability.

Row Sequency	Column Sequency 1	2	3	4	5	6	7	8	Row Marginal Means
1	1.00	2.73	3.18	3.82	6.64	7.64	6.36	5.18	4.57
2	2.27	3.64	6.27	5.09	8.00	9.82	9.27	9.18	6.32
3	4.27	5.36	6.27	7.45	7.82	10.45	9.27	8.27	7.40
4	3.91	4.82	7.55	5.18	8.36	11.36	10.82	5.55	7.19
5	5.55	7.36	8.82	8.45	9.27	11.91	11.64	9.73	9.09
6	7.36	9.91	10.91	10.55	12.27	13.82	13.45	10.55	11.10
7	6.55	8.73	9.18	10.36	12.18	14.18	11.73	10.09	10.38
8	5.09	6.18	7.64	7.09	9.27	11.64	10.09	6.36	7.92
Column Marginal Means	4.50	6.09	7.47	7.25	9.23	11.35	10.33	7.74	8.00

Table 2.3 Mean Complexity Ranks for the Walsh stimuli Ordered by Sequency.

The memorability instruction was used here as a measure of obtaining a measure that is approximately uniform across participants. It conforms to an intuitive notion of what complexity is, though there are other ways of defining complexity which are generally biassed by referring to the physical properties of the stimulus. It will be shown later that a number of the physical features of the Walsh stimuli derived above can be used as estimates of stimulus complexity. The present response-b a sed measure of complexity (based on the memorability criteria) is thus complementary to the stimulus-based measures of complexity.

Overview of the thesis

Analysis

Selection of novel and quantifiable stimuli

↓

Feature selection (physical properties)

↓

Feature extraction (psychological properties)

↓

Feature space representation (MDS and Cluster analysis of similarities data)

↓

Categorical representation

↓

Theory

Summary of previous literature on knowledge representation and utilisation

↓

A framework for studying knowledge acquisition

↓

The use of similarity and sorting tasks to determine how knowledge is represented

↓

The use of modified similarity tasks as tools in developing a psychological epistemology

↓

<u>Conclusion</u>: The analytic methods presented here may be used to identify the knowledge representation of a stimulus set. A set of experimental tasks which is outlined in this thesis should be particularly useful in studying the acquisition of knowledge for a given set of visual stimuli. The Walsh stimuli are not a suitable set of stimuli for studying knowledge acquisition as their logical structure is readily apprehended even in the absence of previous exposure to them.

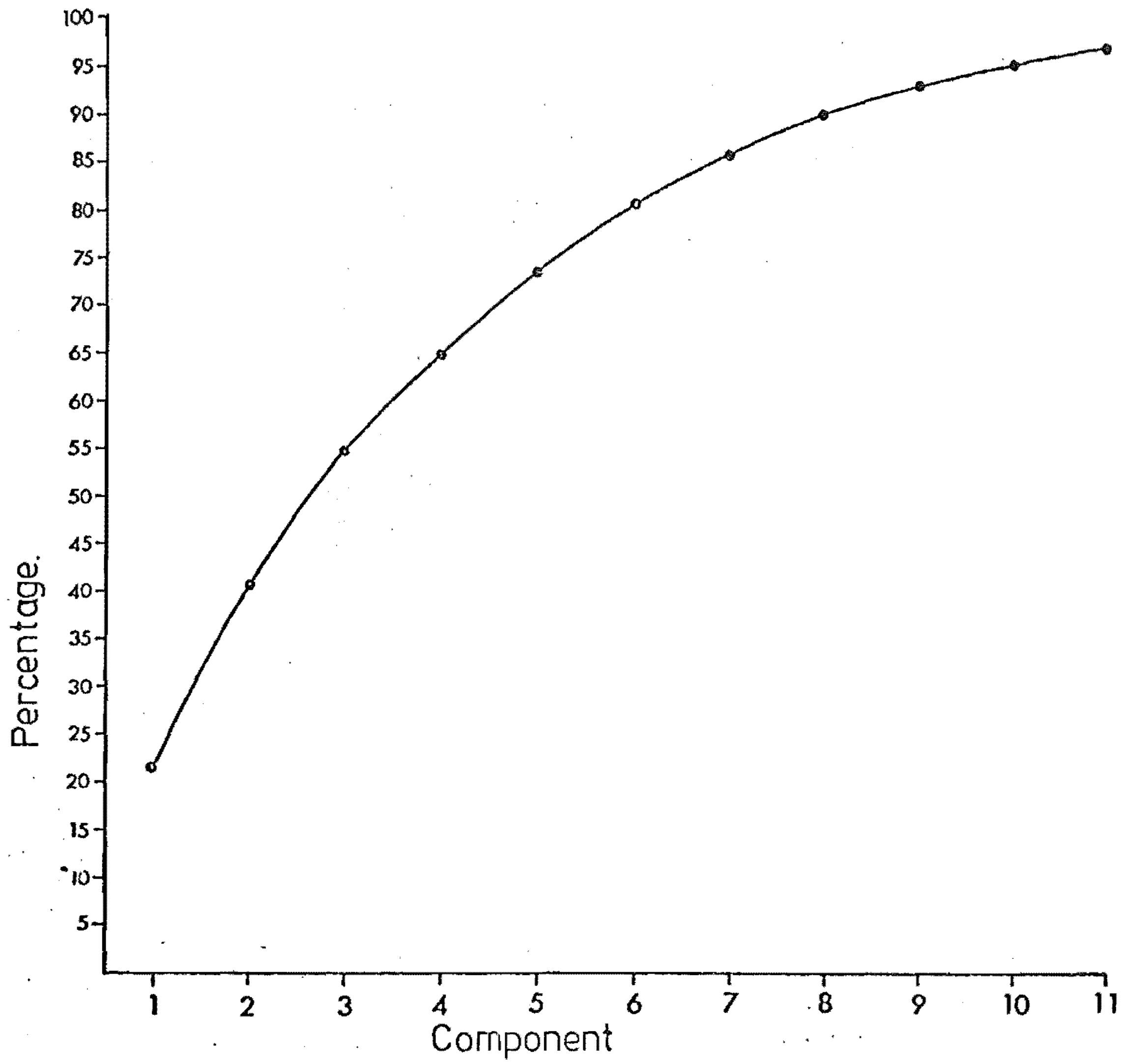


Figure 2.3 Cumulative percent variance accounted for (y-axis) by successive components (x-axis).

Principal Components Analysis of the Walsh Features

The derivation of a set of 23 features based on the physical properties of the Walsh stimuli has been given. In addition, the complexity model described above synthetically generates another feature. The pattern of complexity rankings was approximately the same for all the participants used in E2. In view of this lack of individual variability with respect to it, the empirically derived measure of complexity was added to the 23 features described previously.

A Principal Components Analysis (PCA) was used (BMD01M, Dixon, 1973) on the augmented set of 24 features in an attempt to eliminate descriptive redundancy in the set of features by extracting orthogonal combinations. The feature scores on the 35 stimuli in the upper triangular portion of the matrix (above the negative diagonal) of sequency-order Walsh stimuli (see Figure A.1, Appendix A) were used in the analysis. These 35 stimuli represent all the rotationally distinct Walsh stimuli, out of the complete set of 64. Figure 2.3 shows the relationship between the cumulative proportion of total variance accounted for in the PCA analysis and the number of components. The smooth curve in Figure 2.3 gives no facile graphical indication of how many components should be considered.

The first three component scores for each of the 35 stimuli are given in Table 2.4. An examination of the pattern of component scores across stimuli did not suggest any simple interpretation of the three components.

Stimulus Number	Component 1	Loadings Component 2	Component 3
2	-3.297	-3.060	2.537
3	-3.307	-2.900	0.348
4	-2.869	-2.660	-1.193
5	-2.514	-2.047	-1.679
6	-2.266	-1.453	-1.054
7	-3.172	-0.195	-0.238
8	-5.603	2.142	2.670
10	0.258	-1.906	2.945
11	2.008	-2.605	1.787
12	0.318	-2.625	-0.451
13	1.068	-2.097	-0.607
14	-0.018	-0.802	-0.024
15	-0.117	-0.360	0.397
16	-2.778	0.191	-0.805
19	2.626	-1.540	4.359
20	2.826	-1.494	1.042
21	2.341	-1.220	0.261
22	1.593	-0.267	0.766
23	0.657	-0.178	-0.400
24	-1.751	-0.118	-1.910
28	1.443	-0.384	1.081
29	1.813	0.068	0.089
30	1.888	0.273	-1.211
31	1.378	0.235	-2.160
32	-1.758	0.886	-3.889
37	2.645	0.475	0.201
38	2.384	0.695	-1.134
39	2.115	0.822	-2.176
40	-0.648	1.786	-3.109
46	2.587	1.742	-0.589
47	2.187	2.313	-0.440
48	-0.530	3.029	-1.504
55	2.218	3.185	1.258
56	-0.777	3.998	0.139
64	-2.947	6.073	4.163

Table 2.4 Scores on the First Three Components for each of the Walsh stimuli.

The three components will be used in later chapters along with the other features in attempts to account for differences in responses to different stimuli. The three components are uninterpretable in terms of first-order physical properties and may have arisen solely because of the orthogonal partitioning of the covariance over the feature and stimulus set. The performance of the components in accounting for patterns of responding in experiments will be compared with the performance of the other features, but the components will not be used in building substantive psychometric theories.

Pattern Redundancy

The notion of redundancy has been used extensively in previous studies. One view of redundancy defines it as the amount of information required to define a stimulus (Hochberg and McAlister, 1953). Such a measure of redundancy should be closely related to some of the features already described. It could even be argued that the 64 Walsh masks are indistinguishable in terms of redundancy since they can each be identified by means of two, and only two, parameters (row and column sequency).

A second type of redundancy measure is based on the equivalence set which can be derived from a given stimulus.

The derivation of an equivalence set for each of the Walsh stimuli can be made using the admissible symmetry transformations which do not change the shape of a square (Rosen, 1975). These consist of four unique rotational transformations — 0^{o}, 90^{o}, 180^{o}, 270^{o}, and two mirror reflections (reflections about the horizontal and vertical axes which pass through the centre of the stimulus). Diagonal reflections are also admissable transformations for the preservation of a square shape but they will not be considered here as almost all of the Walsh stimuli are affected by diagonal reflections.

Row Sequency								
8	4	4	4	4	4	4	4	2
7	2	4	2	4	2	4	1	4
6	4	4	4	4	4	2	4	4
5	2	4	2	4	1	4	2	4
4	4	4	4	2	4	4	4	4
3	2	4	1	4	2	4	2	4
2	4	2	4	4	4	4	4	4
1	1	4	2	4	2	4	2	4
Column Sequency	1	2	3	4	5	6	7	8

Table 2.5 Size of Equivalence Sets for each of the Walsh stimuli, Arranged in Sequency Order.

Table 2.5 gives the size of the equivalence set for each of the Walsh stimuli. It can be seen that all the Walsh stimuli have one of three possible equivalence set sizes (1, 2, or 4) and that there is a relationship between sequencies and equivalence set size. Equivalence set size has been used as an estimate of pattern redundance (Glushko, 1975) although the restricted set of permissible transformations for the Walsh stimuli[1] (one, two, or four) argues against equivalence set size being an important feature in perceiving the Walsh stimuli.

The inclusion of equivalence set size makes a total of 28 features of the Walsh stimuli which have been described in this chapter. Table 2.6 gives the measures on these 28 features for each of the 35 non-homogeneous (i.e., excluding the completely black Walsh stimulus) Walsh stimuli which are above or on the negative diag onal (bottom-left to top-right) of the matrix of sequency-ordered Walsh stimuli.

A Notational Point

During the course of this thesis it will be necessary to use different sets of numbers denoting, among other things the Walsh stimuli and the features. To reduce confusion, the

1. Some preliminary analyses were carried out in an attempt to discover a relationship between equivalence set size and preferences as found for other stimulus sets (Garner, 1974) but no relationship was apparent.

Identification of the 28 features in Table 2.6

Feature	
1.	Complexity (as given in Table 2.3)
2.	Column sequency
3.	Row sequency
4.	Block structure
5.	The number of black squares
6.	The number of black rectangles
7.	Average grain
8.	Grain variance
9.	Heterogeneity of area
10.	Heterogeneity of shape
11.	Squareness
12.	Combined heterogeneity.
13*	Mean phi-coefficent (D=1)
14.	Mean entropy (D=1)
15.	Range of the phi-coefficient(D=1)
16.	Range of entropy (D=1)
17.	Mean phi-coefficient (D=2)
18.	Mean entropy (D=2)
19.	Range of the phi-coefficient (D=2)
20.	Range of the entropy (D=2)
21.	Mean phi-coefficient (D=3)
22.	Mean entropy (D=3)
23.	Range of phi-coefficient (D=3)
24.	Range of entropy (D=3)
25.	Component one (as given in Table 2.4)
26.	Component two
27.	Component three
28.	Equivalence set size (as given in Table 2.5)

*Features 13-24 are the 12 textural features whose derivation is outlined in Appendix B.

TABLE 2.6 (Part 1): 35 of the Walsh Stimuli Quantified in terms of 28 Features.

Stimulus Number	Features 1	2	3	4	5	6	7	8	9	10	11	12
2	2.27	1	2	1	0	1	24.0	.00	.00	.00	.00	.00
3	4.27	1	3	1	0	1	24.0	.00	.00	.00	.00	.00
4	3.91	1	4	3	0	2	20.0	.00	.00	.00	.00	.00
5	5.55	1	5	3	0	3	18.7	.89	.00	.00	.00	.00
6	7.36	1	6	1	0	3	18.7	.89	.00	.00	.00	.00
7	6.55	1	7	1	0	3	18.7	.89	.00	.00	.00	.00
8	5.09	1	8	3	0	4	18.0	.00	.00	.00	.00	.00
10	3.64	2	2	2	2	0	16.0	.00	.00	.00	1.00	.00
11	5.36	2	3	2	1	2	13.3	3.56	16.00	1.00	.33	16.00
12	4.82	2	4	1	0	4	12.0	.00	.00	.00	.00	.00
13	7.36	2	5	1	0	5	11.2	.96	8.00	.50	.00	4.00
14	9.91	2	6	2	0	6	10.7	.89	12.80	.00	.00	.00
15	8.73	2	7	2	0	7	10.3	.49	5.30	.00	.00	.00
16	6.18	2	8	1	0	8	10.0	.00	.00	.00	.00	.00
19	6.27	3	3	2	5	0	9.6	10.24	14.40	.00	1.00	.00
20	7.55	3	4	3	4	2	9.3	3.56	16.00	1.00	.67	16.00
21	8.82	3	5	1	4	3	9.1	2.12	8.00	1.00	.57	8.00
22	10.91	3	6	2	2	7	8.0	4.44	15.20	.40	.22	6.08
23	9.18	3	7	2	0	11	7.3	4.56	14.70	.00	.00	.00
24	7.64	3	8	1	0	12	7.3	3.56	4.00	.00	.00	.00
28	5.18	4	4	3	8	0	8.0	.00	.00	.00	1.00	.00
29	8.45	4	5	3	6	4	7.2	.96	2.00	.50	.60	1.00
30	10.55	4	6	1	4	8	6.7	.89	3.20	.75	.33	2.40
31	10.36	4	7	1	2	12	6.3	.49	.70	.33	.14	2.30
32	7.09	4	8	3	0	16	6.0	.00	.00	.00	.00	.00
37	9.27	5	5	3	9	4	6.2	2.75	4.25	.50	.69	2.13
38	12.27	5	6	1	7	8	5.9	1.85	3.50	.50	.47	1.80
39	12.18	5	7	1	5	12	5.6	1.05	1.00	.50	.29	.50
40	9.27	5	8	3	8	12	5.2	.96	.50	.50	.40	.25
46	13.82	6	6	2	10	8	5.3	1.78	3.20	.32	.56	1.02
47	14.18	6	7	2	13	8	5.0	1.38	1.80	.60	.62	1.08
48	11.64	6	8	1	16	8	4.7	.89	.80	.50	.67	.40
55	11.73	7	7	2	21	4	4.5	1.05	1.39	.39	.84	.54
56	10.09	7	8	1	24	4	4.3	.49	.33	.33	.86	.11
64	6.36	8	8	3	32	0	4.0	.00	.00	.00	1.00	.00

TABLE 2.6 (Part 2): 35 of the Walsh Stimuli Quantified in terms of 28 Features.

Stimulus Number	Features 13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28
2	.812	1.332	.249	.447	.625	1.434	.499	.583	.437	1.274	.749	.369	-3.297	-3.060	2.537	4
3	.562	1.626	.583	.839	-.125	1.434	1.500	.583	-.250	1.387	1.666	.520	-3.307	-2.900	0.348	2
4	.347	1.707	.833	.947	-.500	.935	2.000	.081	.125	1.686	1.166	.919	-2.869	-2.660	-1.193	4
5	.124	1.707	1.116	.947	-.500	.997	2.000	.000	.374	1.686	.833	.919	-2.514	-2.047	-1.679	2
6	-.063	1.626	1.416	.839	-.001	1.683	1.333	.915	.750	1.387	.333	.520	-2.266	-1.453	-1.054	4
7	-.313	1.332	1.750	.447	.500	1.683	.666	.915	.374	1.686	.833	.919	-3.172	-0.195	-0.238	2
8	-.500	.986	2.000	.014	1.000	.997	.000	.000	-.500	.975	2.000	.029	-5.603	2.142	2.670	4
10	.612	1.691	.204	.211	.216	1.940	.233	.056	-.152	1.880	.097	.170	0.258	-1.906	2.945	2
11	.439	1.823	.407	.338	-.050	1.940	.666	.056	-.132	1.900	.736	.249	2.008	-2.605	1.787	4
12	.265	1.884	.612	.400	-.342	1.670	1.333	.915	-.088	1.969	.225	.007	0.318	-2.625	-0.451	4
13	.092	1.884	.857	.400	-.334	1.683	1.333	.915	-.013	1.969	.400	.067	1.068	-2.097	-0.607	4
14	-.082	1.823	1.142	.338	-.056	1.948	.666	.072	.031	1.900	.800	.249	-0.018	-0.802	-0.024	4
15	-.255	1.691	1.428	.211	.222	1.948	.222	.072	-.013	1.969	.400	.007	-0.117	-0.360	0.397	4
16	-.429	1.437	1.714	.588	.500	1.683	.666	.915	-.206	1.714	1.188	.968	-2.778	0.191	-0.805	4
19	.305	1.913	.245	.111	-.117	1.940	.433	.056	-.121	1.806	.958	.180	2.626	-1.540	4.359	1
20	.173	1.954	.366	.133	-.159	1.670	1.350	.915	-.140	1.910	.721	.265	2.826	-1.494	1.042	4
21	.040	1.954	.571	.133	-.167	1.683	1.333	.915	-.161	1.910	.800	.265	2.341	-1.220	0.261	2
22	-.092	1.913	.857	.111	-.112	1.948	.444	.072	-.180	1.806	1.199	.180	1.593	-0.267	0.766	4
23	-.225	1.823	1.142	.338	-.056	1.948	.666	.072	-.161	1.910	.800	.265	0.657	-0.178	-0.400	2
24	-.358	1.642	1.428	.860	-.001	1.683	1.333	.915	-.099	1.533	1.602	.719	-1.751	-0.118	-1.910	4
28	.081	1.986	.122	.013	.000	.992	2.000	.008	-.081	1.978	.238	.025	1.443	-0.384	1.081	2
29	-.010	1.986	.285	.013	.000	.997	1.000	.000	-.020	1.978	.400	.025	1.813	0.068	0.089	4
30	-.103	1.954	.571	.133	-.167	1.683	1.333	.915	.039	1.910	.800	.265	1.888	0.273	-1.211	4
31	-.194	1.884	.857	.400	-.334	1.683	1.333	.915	-.020	1.978	.400	.025	1.378	0.235	-2.160	4
32	-.286	1.733	1.142	.982	-.500	.997	2.000	.000	-.201	1.722	1.198	.968	-1.758	0.886	-3.889	4
37	-.062	1.986	.162	.013	.000	.997	2.000	.000	.119	1.978	.161	.025	2.645	0.475	0.201	1
38	-.113	1.954	.490	.133	-.167	1.683	1.333	.915	.260	1.910	.478	.265	2.384	0.695	-1.134	4
39	-.164	1.884	.815	.400	-.334	1.683	1.333	.915	.119	1.978	.161	.025	2.115	0.822	-2.176	2
40	-.215	1.733	1.143	.982	-.500	.997	2.000	.000	-.300	1.722	1.200	.968	-0.648	1.786	-3.109	4
46	-.123	1.913	.611	.111	-.112	1.948	.444	.072	.479	1.806	.241	.180	2.587	1.742	-0.589	2
47	-.133	1.823	1.020	.338	-.056	1.948	.666	.072	.260	1.910	.478	.265	2.187	2.313	-0.440	4
48	-.143	1.642	1.428	.860	-.001	1.683	1.333	.915	-.402	1.533	1.600	.719	-0.530	3.029	-1.504	4
55	-.103	1.691	1.224	.211	.222	1.948	.222	.072	.119	1.978	.161	.025	2.218	3.185	1.258	1
56	-.072	1.437	1.715	.588	.500	1.683	.666	.915	-.300	1.722	1.200	.968	-0.777	3.998	0.139	4
64	.000	.996	2.000	.000	1.000	.997	.000	.000	.000	.996	2.000	.001	-2.947	6.073	4.163	2

features will be numbered[1] as in the identification chart (p.28) and will be described as feature one, feature two etc. The Walsh stimuli will be numbered according to their position in the sequency-ordered matrix (Figure A.1, Appendix A), being numbered in successive columns of this matrix. Thus the stimuli in the first column are numbered one to eight from the bottom to the top of that column. The stimuli in the second column are numbered nine to sixteen, and so on. The stimuli will be referred to as stimulus #1, stimulus #2, etc. where the number refers to the ordered position in the sequency-ordered matrix as explained above.

Summary

A total of 28 features of the Walsh stimuli have now been outlined. Of these features, twelve (the first eleven features plus the equivalence set size) are interpretable in terms of first-order physical properties of the Walsh stimuli. The twelve textural features are not so easily explained in terms of simple visual properties of the Walsh stimuli. The three sets of component scores may be regarded as useful features for explaining variation in responses although they will not be used in psychological theory building. Table 2.7 shows the matrix of intercorrelations for 27 of the features. It can be seen that the three components each have at least one moderately large correlation (abs (r) >.7) with one of the other 24 features (equivalence set size was not included in the table). One other feature which may be important in the perception of the Walsh stimuli is preference. This feature should be important for both theoretical and practical reasons, which will be outlined in the next chapter, where six scales of preference for the Walsh stimuli are derived.

1. Some preliminary analyses were carried out in an attempt to discover a relationship between equivalence set size and preference as found for other stimulus sets (Garner, 1974), but no relationship was apparent.

Variable Number	Features 1	2	3	4	5	6	7	8	9	10	11
1	1.000										
2	0.369	1.000									
3	-0.037	0.000	1.000								
4	0.558	0.346	0.136	1.000							
5	0.424	0.667	-0.122	-0.136	1.000						
6	-0.782	-0.551	-0.098	-0.573	-0.516	1.000					
7	0.097	-0.190	-0.003	-0.073	-0.040	-0.205	1.000				
8	0.008	-0.276	0.032	-0.012	-0.224	-0.074	0.847	1.000			
9	0.448	-0.053	-0.110	0.184	0.080	-0.406	0.193	-0.038	1.000		
10	0.527	-0.128	0.230	0.729	-0.361	-0.498	0.239	0.309	0.316	1.000	
11	-0.030	-0.305	-0.027	-0.103	-0.166	-0.060	0.317	0.070	0.762	0.119	1.000
12	-0.325	-0.907	-0.060	-0.110	-0.652	0.535	0.049	0.181	0.008	0.167	0.214
13	0.391	-0.330	0.002	-0.238	0.201	-0.344	0.363	0.203	0.475	0.150	0.324
14	-0.110	0.800	-0.047	0.246	0.257	-0.061	-0.283	-0.247	-0.347	-0.292	-0.338
15	-0.123	0.188	-0.150	-0.229	0.326	0.292	-0.247	-0.222	-0.279	-0.488	-0.258
16	-0.201	0.193	-0.055	0.365	-0.378	0.133	-0.183	-0.081	-0.277	0.179	-0.169
17	0.182	-0.046	-0.615	-0.076	0.099	-0.155	0.382	0.287	0.195	0.021	0.229
18	0.105	-0.102	0.218	-0.233	0.207	-0.017	-0.155	-0.226	0.169	-0.136	0.013
19	0.019	0.110	-0.738	-0.114	0.139	0.025	-0.126	-0.170	0.226	-0.221	0.126
20	-0.099	-0.176	-0.124	-0.071	-0.131	0.246	-0.084	-0.110	-0.112	-0.146	-0.148
21	0.444	-0.178	-0.048	-0.156	0.221	-0.416	0.221	0.123	0.420	0.162	0.263
22	-0.274	0.358	0.077	0.147	0.032	0.122	0.001	0.019	-0.213	-0.157	-0.087
23	-0.051	0.272	-0.087	-0.101	0.220	0.204	-0.151	-0.152	-0.203	-0.293	-0.185
24	0.601	-0.199	-0.033	0.162	0.112	-0.613	0.465	0.306	0.657	0.500	0.439
25	0.579	0.766	0.162	0.824	0.289	-0.680	-0.151	-0.148	0.075	0.451	-0.258
26	-0.297	-0.452	0.129	0.256	-0.774	0.224	0.293	0.442	-0.103	0.471	0.153
27	0.840	0.447	-0.092	0.501	0.466	-0.770	0.154	-0.018	0.428	0.365	-0.024

Variable Number	Features 12	13	14	15	16	17	18	19	20	21	22	23
12	1.000											
13	0.121	1.000										
14	-0.638	-0.726	1.000									
15	0.007	-0.163	0.321	1.000								
16	-0.111	-0.797	0.467	-0.370	1.000							
17	0.088	0.249	-0.044	-0.201	-0.012	1.000						
18	0.036	0.495	-0.349	0.428	-0.696	-0.574	1.000					
19	-0.085	0.017	0.099	0.197	-0.043	0.280	0.127	1.000				
20	0.139	0.093	-0.120	0.023	-0.085	0.092	0.041	0.065	1.000			
21	-0.017	0.853	-0.523	-0.236	-0.603	0.380	0.275	0.027	0.044	1.000		
22	-0.176	-0.637	0.562	0.320	0.356	-0.317	-0.135	-0.021	-0.508	-0.752	1.000	
23	-0.130	-0.272	0.390	0.812	-0.127	-0.228	0.307	0.175	-0.029	-0.249	0.438	1.000
24	0.043	0.834	-0.619	-0.513	-0.451	0.437	0.097	-0.010	0.050	0.810	-0.645	-0.514
25	-0.595	-0.367	0.559	-0.168	0.420	-0.101	-0.280	-0.117	-0.180	-0.236	0.294	-0.003
26	0.447	-0.327	-0.145	-0.621	0.638	0.131	-0.644	-0.311	-0.098	-0.289	0.095	-0.483
27	-0.441	0.383	-0.001	-0.279	-0.166	0.378	-0.107	-0.013	0.162	0.410	-0.311	-0.236

Variable Number	Features 24	25	26	27
24	1.000			
25	-0.000	1.000		
26	-0.009	-0.008	1.000	
27	0.635	0.592	-0.262	1.000

Table 2.7 Intercorrelation matrix for the first 27 features outlined in this chapter.

CHAPTER III

In chapter one there was outlined an overall framework for knowledge and its acquisition using as an explanatory concept the notion of the person's increasingly adaptive fitting to the environment. This notion has previously been used in adaptation level theory:

"(A)n individual's attitudes, values, ways of structuring his experience, judgements of physical, aesthetic and symbolic objects, intellectual and emotional behaviour, learning, interpersonal relations all represent modes of adaptation to environmental and organismic forces."

— Helson (1964, p.37)

As a consequence, adaptation level theorists have viewed sensory input as being analysed in terms of the person's preceding and concurrent sensory processes, memory variables, and physiological attributes; the perceiver having a continuing dynamic interaction with the environment (Murch, 1973, p.258). I shall also use this perspective in the present thesis, though with experimental paradigms which are different from the psychophysical tasks generally used in previous studies.

It seems likely that part of the process of adaptive fitting to the environment should involve the evaluation of perceptual input in terms of its desirability or usefulness. A person's preference for one type of perceptual input as against another indicates the relative desirability of that input. This chapter will attempt to use preference judgements as a way of investigating the role of relative desirability of perceptual inputs in perceptual learning. I shall first consider what exactly is involved in making a preference judgement before attempting to interpret experimental results that are based on such preference judgements. The following quotation introduces the salient issues:

"The word preference, in normal language, usually refers to the desirability of some object or event . . . The essential features of an act of preference appears to be the selection of this one, rather than some other, response . . . In general, it is the ends of actions that are preferred . . . Of course actions that we take in order to reach valued outcomes are themselves often objects of value and therefore objects of preference. Often, a particular stimulus object, such as a Van Gogh painting, may be an object of preference, but the fact that we 'look at' the painting already suggests that it is the outcome quality of the act of looking that is preferred."

— Galanter (1966, pp.55–56).

It appears that Galanter is implying a hierarchy of preference objects, starting with the valued outcomes themselves and working back through the actions necessary to reach those outcomes. Thus, an object would originally be preferred if it lead directly to a valued outcome, while objects which were only indirectly related to the valued outcome would become objects of preference if they were incorporated into a response chain (Keller and Schoenfeld, 1950, pp. 197–208), which would eventually lead (at least in principle) to the valued outcome. This notion of the relationship between an object and a valued outcome appears analogous to Gibson's (e.g. 1977) concept of the 'affordance' of an object to a perceiver.

It is difficult to see how the Walsh stimuli can be even indirectly related to valued outcomes, since few of them are part of the average person's environment. Galanter (in the quotation above) suggests that it might be the outcome quality of the act of looking that is preferred, but it is hard to conceptualize what the outcome quality of a Walsh stimulus would be.

This difficulty in explaining why preference judgements can be made with respect to stimuli that have no apparent affordance structure (not even of the indirect kind mentioned by Galanter) reflects the lack of a clear definition of what preference is, vis a vis aesthetic perception (Berlyne, 1971), hedonic judgement (Young, 1952) preferential choice (Coombs, 1964) and affectivity (Helson, 1964). One way of resolving this difficulty is to distinguish between the functional notion of affordance, and the more abstract concept of aesthetic perception where the beauty of an object is judged independently of the affordance of that object to the perceiver (this distinction has been made previously by Morris, 1956). Since the Walsh stimuli can be expected to have little in the way of affordance structure for most perceivers I shall assume that the preference judgements elicited in Experiment E3 (described below) reflect the aesthetic perception of the participants, rather than the functional pleasingness or affordance which the stimuli have for them. Using a similar set of black and white checkerboard stimuli to the ones used here, Smets (1973) found that two rating scales which were apparently designed to measure the affordance structure and aesthetic value of the stimuli, were in fact only measuring what I assume to be the aesthetic perception of the participants.

Smets' participants rated the set of stimuli on two bipolar nine-point scales:

1. pleasant vs unpleasant.
2. ugly vs beautiful.

Smets found that the scale pleasant-unpleasant provided greater differentiation (in terms of redundancy) than the ugly-beautiful scale, although the two scales appeared to give roughly similar results.

Preference and Perceptual Processing

In agreement with the present distinction between affordance structure and aesthetic value, Berlyne (1971) has claimed that aesthetic perception is *not* purposive and functional. Within the theoretical framework of this thesis, however, it is necessary to discover what, if any, effect aesthetic perception, or more generally, preference, has on perceptual processing.

Helson (1964, pp. 337–342) has outlined some of the early evidence for a relationship between perception and 'affectivity' and concluded that affectivity and perception interact, without attempting to decide which was the cause and which the effect. Many studies have tended to view preference as an effect rather than a cause (see the discussion on preference and complexity below), few attempts have been made to investigate the role of preference in influencing subsequent perceptual processing. The effect of preference for the stimuli in a pair on a pair-wise similarity judgement will be considered in chapter seven. It appears that preference may have an effect on reaction time (Shipley, Coffin, and Hadsell, 1945; Hoosain, 1977) but that possibility will not be investigated in this thesis.

Individual Differences in Preferences

The old adage holds that "Beauty is in the eye of the beholder." A more sophisticated discussion of this subjectivist view of aesthetic perception is given by Gregson (1961). If there are individual differences in preference judgements, then they need to be accounted for before the relationship between preference and perceptual processing can be properly characterised. A study of the previous literature indicates evidence both for and against the existence of individual differences in preference responding. Mavrides (1970) concluded that, for her patterns of symmetrical star-shaped outlines, "preferences are not completely an individual phenomenon but are held by (participants) generally". However, Mavrides pooled judgements across her participants and based the idea of a single preference scale on the homogeneity of the pooled results between the different experimental conditions. The fact that pooled estimates are robust across experimental tasks does not, however, exclude the possibility that a consistent set of homogeneous subgroups underlie the pooled results.

Valentine (1962, p.99) suggested the following three types of criteria for rating preference in geometrical forms:

1. formal criteria, such as symmetry, complexity and organisation,
2. connotative criteria such as familiarity,
3. potential for design,

of which only the first (formal criteria) would be used by the majority of people. This appears to follow the more extensive work of Birkhoff (1929, 1933) which is now not readily available (Gregson, personal communication). Davis (1936) used a preference ranking experiment with visual stimuli (polygons) and analysed the results separately for art students and non-art students. His results indicated that there was "no significant difference between the preferences of art students and those of students in general." Eysenck and Castle (1970) in a similar experiment found little difference between the art and non-art students except that the artists tended to prefer simple polygons while the non-artists preferred complex polygons.

The differences between those who are interested in art, and those who are not, is only one possible way of splitting people into groups which may or may not differ in terms of their preferences for individual forms. It would seem more sensible to first detect what individual differences there are in aesthetic perception before attempting to relate types or patterns of aesthetic perception to attributes that people have.

Individual differences may well be an important factor in preference judgements (note the experiment E3 results presented later in this chapter), but this possibility seems to have been ignored by a large portion of the relevant literature. As a recent example of this, Cupchik and Berlyne (1979) pooled results and used analysis of variance and t-tests in relating complexity preference. While such pooling of results *may* be valid in determining overall trends, it is necessary to account for individual differences in modelling the relationship between preference and perceptual processing, as shown below. (See also Woodworth, 1938, Chapter 16).

An experiment (E3, described below) was carried out to determine the preferences of individuals for the Walsh stimuli. In analysing the results the emphasis was placed on identifying homogeneous subgroups of participants which differed in terms of their implied preference scales.

Experiment E3

Experiment E3 sought to scale the Walsh stimuli in terms of preference. The discussion above suggests that:

1. The preference instructions will elicit judgements based on aesthetic perception, and

2. there may actually be several preference scales attributable to the Walsh stimuli, used by different clusters of individuals or by the same individual at different times.

Method

44 participants were each asked to construct a two-way (7 x 5) conceptual ranking of the Walsh stimuli with respect to their preference for each of the stimuli. The conceptual ranking method was the same as that used in experiment E2 except for the following changes:

1. The rows and columns were ranked according to preference, instead of the constrained measure of complexity used in E2.

2. Only 35 (instead of all 64) of the Walsh stimuli were used. These 35 were selected in the following manner:

(a) 28 stimuli in the portion of the sequency ordered matrix (see Table A.1 in Appendix A) below the bottom left top-right (negative) diagonal were removed as they were simply 90^{o} rotations of the Walsh stimuli opposite them in the matrix.

(b) Out of the remaining 36 stimuli the homogeneous black stimulus in the bottom left-hand corner was removed. (This stimulus is atypical in that it is the only one that does not contain both black and white).

3. The remaining 35 stimuli were arranged into five rows and seven columns which were ranked in the following order: [1]

Order of Ranking

1 column one

2 row one

3 column two

4 row two

5 column three

6 column four

7 row three

8 column five

9 row four

10 column six

11 row five

12 column seven

Tables 3.1 and 3.2 show the estimated means and standard deviations of each of the ranks assigned to each of the positions in the conceptual grid. (100 simulation runs were used to obtain these estimates). Table 3.1 was used to convert the results of E3 into implied preference ranks for each of the Walsh stimuli over each of the 44 participants. These ranks were then used as estimated preference scale values in the following analysis.

Row	Column 1	2	3	4	5	6	7
1	1.00	2.41	4.39	7.03	10.63	16.07	25.09
2	3.08	5.88	8.96	12.05	16.10	21.42	29.14
3	5.75	10.48	13.48	16.19	19.55	24.67	31.71
4	9.32	15.35	18.88	21.73	24.15	27.72	33.65
5	16.61	23.01	26.70	29.06	31.01	32.73	35.00

Table 3.1 Estimated ranks for each conceptual grid position (obtained from 100 simulation trials).

1. Monte Carlo simulation was run to see what the expected ranks for each position in the conceptual grid would be after this procedure was carried out.

Row	Column						
	1	2	3	4	5	6	7
1	0.000	0.618	1.264	1.957	2.671	3.623	3.496
2	0.987	1.373	1.549	1.862	2.256	2.499	1.944
3	1.785	1.982	1.763	1.765	1.734	1.919	1.388
4	2.756	2.886	2.511	1.943	1.883	1.692	0.740
5	4.743	3.183	1.988	1.495	1.212	0.989	0.000

Table 3.2 Estimated standard deviations in ranks for each conceptual grid position.

Results and Analysis

The intercorrelations between the 44 implied preference scales are given in Table 3.3. This intercorrelation matrix was then analysed using average weighted hierarchical clustering (Sokal and Sneath, 1963), which was implemented using BMDP1M (Dixon, 1975). Figure 3.1 shows the results of the cluster analysis printed in the form of a tree. The accompanying comments indicate how the tree is to be interpreted.

On the basis of Figure 3.1 I selected a total of six clusters involving 27, 5, 3, 2, 2, and 2 cases respectively (accounting for a total of 41 of the 44 cases). Clusters two, four, five and six, each had average intercorrelations of greater than .5 between their members while clusters one and three had average intercorrelations of about .45 between their members. The preference scales (the implied preference rankings for each of the 35 stimuli) were separately pooled and averaged for each of the six clusters to obtain six preference scales which are given in Table 3.4.

Stimulus Number	Preference Clusters 1	2	3	4	5	6
2	3.43	12.86	30.97	21.10	8.20	2.40
3	5.18	10.98	29.80	22.40	8.20	14.55
4	7.23	16.06	27.93	10.30	18.95	14.10
5	8.36	8.86	30.37	22.35	28.85	19.00
6	6.95	4.82	24.40	27.60	26.30	22.95
7	12.06	8.98	29.23	18.75	29.40	29.35
8	11.77	21.22	23.97	12.75	33.35	22.15
10	10.49	27.60	27.23	13.35	23.10	13.80
11	8.01	13.90	22.60	29.20	12.90	3.40
12	13.50	24.98	22.53	5.80	13.80	8.00
13	12.13	11.02	23.37	22.80	6.40	9.25
14	12.85	5.50	20.30	31.35	2.75	10.95
15	14.14	9.14	22.13	30.40	10.25	17.50
16	13.67	23.14	22.90	8.95	27.10	13.35
19	18.44	30.56	27.60	9.25	16.00	15.85
20	17.04	15.70	18.63	22.85	12.75	3.75
21	18.63	20.26	17.07	17.55	14.10	16.15
22	17.57	4.68	17.93	29.85	9.60	24.85
23	23.14	14.26	18.70	18.10	15.05	26.40
24	18.70	19.40	21.33	15.05	15.60	19.20
28	20.04	33.76	15.00	8.95	28.85	11.25
29	18.99	27.48	12.40	5.85	12.70	8.20
30	21.53	11.44	8.97	27.90	5.00	19.00
31	21.46	11.18	4.17	24.05	11.30	27.95
32	22.80	28.28	14.53	11.20	32.20	18.60
37	23.54	28.24	13.00	5.70	23.05	8.75
38	26.61	18.46	10.43	27.05	5.85	24.85
39	29.50	15.28	5.33	9.80	24.00	29.10
40	24.08	23.38	11.83	6.80	26.20	13.35
46	31.26	16.96	23.47	27.70	12.05	35.00
47	28.22	12.92	8.03	29.35	16.00	29.20
48	26.09	18.44	4.20	16.10	19.60	25.30
55	31.83	28.06	5.33	20.60	22.15	32.70
56	26.83	20.98	6.27	16.05	26.65	20.60
64	24.31	31.52	8.33	3.45	32.05	9.70

Table 3.4 Mean implied preference rankings for the 35 stimuli, pooled within each of

		X(1)	X(2)	X(3)	X(4)	X(5)	X(6)	X(7)	X(8)	X(9)	X(10)
		1	2	3	4	5	6	7	8	9	10
X(1)	1	1.0000									
X(2)	2	0.4752	1.0000								
X(3)	3	0.5149	0.5013	1.0000							
X(4)	4	0.7935	0.3978	0.2737	1.0000						
X(5)	5	0.7308	0.3551	0.2068	0.6085	1.0000					
X(6)	6	0.8289	0.6663	0.7296	0.6387	0.5108	1.0000				
X(7)	7	0.7055	0.2870	0.1970	0.5726	0.6097	0.5117	1.0000			
X(8)	8	0.8776	0.4500	0.5940	0.7043	0.5278	0.8622	0.5131	1.0000		
X(9)	9	0.4254	-0.0345	-0.0559	0.2215	0.4021	0.0954	0.4640	0.1738	1.0000	
X(10)	10	0.6714	0.6847	0.5136	0.4715	0.5738	0.6976	0.5962	0.5304	0.2991	1.0000
X(11)	11	-0.1727	-0.5248	-0.4870	-0.2749	0.0357	-0.4477	0.0398	-0.3349	0.5518	-0.1471
X(12)	12	0.7683	0.3255	0.3602	0.7439	0.4731	0.6966	0.4227	0.7867	0.2050	0.4205
X(13)	13	0.5192	0.7831	0.7620	0.3249	0.3378	0.7729	0.2885	0.5692	-0.1389	0.6297
X(14)	14	0.2965	0.1354	-0.0667	0.3249	0.3000	0.0311	0.3092	0.1212	0.4798	0.1395
X(15)	15	0.5065	0.1030	-0.0127	0.4480	0.5500	0.3206	0.3030	0.3797	0.4980	0.2457
X(16)	16	0.4198	-0.2892	-0.0453	0.3105	0.3207	0.0901	0.3773	0.3750	0.4744	0.1158
X(17)	17	-0.1027	0.1164	0.1745	-0.1168	0.0011	0.0323	-0.1095	-0.0125	0.0732	-0.1000
X(18)	18	0.5845	-0.1046	0.2278	0.4377	0.4497	0.3386	0.4110	0.5269	0.5370	0.3768
X(19)	19	0.8114	0.6369	0.6430	0.7154	0.5155	0.8526	0.5911	0.7734	0.1775	0.7794
X(20)	20	0.8095	0.5520	0.4577	0.7362	0.6221	0.7365	0.5658	0.7154	0.3476	0.7562
X(21)	21	0.8733	0.4675	0.5667	0.7272	0.5208	0.8862	0.5172	0.8790	0.1403	0.5996
X(22)	22	-0.5036	0.1319	0.0705	-0.3690	-0.5603	-0.1243	-0.4310	-0.3134	-0.6443	-0.1914
X(23)	23	-0.5820	-0.4869	-0.3862	-0.6074	-0.2817	-0.6536	-0.3807	-0.5599	0.0720	-0.4149
X(24)	24	0.8062	0.3159	0.4142	0.5587	0.5034	0.7586	0.4667	0.8447	0.3353	0.4916
X(25)	25	0.7670	0.2987	0.3332	0.6394	0.4900	0.6296	0.5704	0.7029	0.3676	0.6047
X(26)	26	0.7524	0.2926	0.4079	0.6272	0.6917	0.6117	0.6009	0.6530	0.5988	0.6386
X(27)	27	0.5391	-0.1229	0.0131	0.3583	0.3994	0.2693	0.4185	0.5012	0.5823	0.2249
X(28)	28	-0.4976	-0.3575	-0.3405	-0.3906	-0.2746	-0.5700	-0.5344	-0.5259	-0.0532	-0.5453
X(29)	29	0.4791	0.5713	0.5104	0.4550	0.4374	0.6371	0.1278	0.4626	-0.0718	0.3217
X(30)	30	0.6392	0.4446	0.4325	0.4521	0.5157	0.5078	0.5351	0.5027	0.5545	0.7429
X(31)	31	0.6825	0.5202	0.5720	0.7244	0.5542	0.8814	0.6169	0.8701	0.2669	0.7116
X(32)	32	0.6739	0.6005	0.6011	0.7307	0.6791	0.8711	0.6661	0.8231	0.2758	0.8453
X(33)	33	0.7689	0.5333	0.6542	0.6037	0.5167	0.7819	0.5500	0.7573	0.2315	0.6623
X(34)	34	0.8102	0.6314	0.5265	0.6933	0.6702	0.8225	0.6585	0.7390	0.1666	0.7790
X(35)	35	0.3539	0.4865	0.3004	0.3934	0.5402	0.4249	0.3435	0.2965	-0.0968	0.3929
X(36)	36	-0.1133	0.5453	0.2469	0.0531	-0.0405	0.1931	-0.0210	-0.0675	-0.3930	0.2713
X(37)	37	0.5749	0.5649	0.5376	0.4243	0.4473	0.6603	0.3773	0.5077	0.1141	0.6205
X(38)	38	0.3505	0.7455	0.6158	0.2655	0.2516	0.5224	0.3575	0.2953	-0.0237	0.6747
X(39)	39	0.1916	-0.0545	-0.2932	0.2032	0.2241	-0.0346	0.1438	0.1240	0.4617	-0.1827
X(40)	40	0.2275	-0.2572	-0.3039	0.1061	0.3308	-0.1387	0.3694	0.0489	0.5282	0.1133
X(41)	41	0.3507	-0.0945	-0.1515	0.3097	0.3245	0.1659	0.0935	0.3647	0.2590	-0.1576
X(42)	42	0.7330	0.6896	0.6851	0.6110	0.3675	0.7905	0.4952	0.7613	0.0794	0.7167
X(43)	43	-0.7473	-0.5285	-0.5979	-0.5230	-0.4065	-0.7922	-0.3879	-0.8183	-0.1997	-0.6661
X(44)	44	0.4993	0.6512	0.7441	0.4050	0.3672	0.7479	0.3488	0.4946	-0.1172	0.6403

Table 3.3 Intercorrelation matrix of the 44 individual preference scales (This table is continued on the next two pages).

		X(11) 11	X(12) 12	X(13) 13	X(14) 14	X(15) 15	X(16) 16	X(17) 17	X(18) 18	X(19) 19	X(20) 20
X(11)	11	1.0000									
X(12)	12	-0.2929	1.0000								
X(13)	13	-0.4853	0.3230	1.0000							
X(14)	14	0.1356	0.2813	-0.0767	1.0000						
X(15)	15	-0.0464	0.3988	0.0263	0.2347	1.0000					
X(16)	16	0.3826	0.4195	-0.2567	0.3419	0.3563	1.0000				
X(17)	17	0.0301	-0.3364	0.1093	0.0165	0.0639	-0.1263	1.0000			
X(18)	18	0.2388	0.5534	0.0139	0.1885	0.3708	0.7055	-0.0432	1.0000		
X(19)	19	-0.3492	0.7020	0.7168	0.2214	0.1711	0.1550	-0.1490	0.3831	1.0000	
X(20)	20	-0.2213	0.7627	0.4823	0.3209	0.3641	0.3531	-0.1469	0.5393	0.8534	1.0000
X(21)	21	-0.3362	0.8340	0.5506	0.0838	0.3194	0.2931	-0.1780	0.5034	0.8314	0.7925
X(22)	22	-0.2954	-0.4365	0.1236	-0.4718	-0.4874	-0.6063	0.1717	-0.6438	-0.1801	-0.3679
X(23)	23	0.5134	-0.7151	-0.4795	-0.1163	-0.1017	-0.0590	0.2207	-0.1058	-0.6644	-0.5684
X(24)	24	-0.0707	0.7095	0.4843	0.0162	0.4022	0.3988	-0.1167	0.5740	0.6423	0.6240
X(25)	25	-0.0369	0.8158	0.3275	0.1899	0.3399	0.5151	-0.3058	0.6302	0.7113	0.8365
X(26)	26	-0.0224	0.5595	0.3295	0.2449	0.6935	0.3949	0.0770	0.6675	0.5936	0.7276
X(27)	27	0.3112	0.5416	-0.0384	0.1880	0.4130	0.5922	-0.0719	0.7591	0.3100	0.4032
X(28)	28	0.0552	-0.4655	-0.3843	0.0505	0.1609	-0.2823	0.2229	-0.2876	-0.6577	-0.5691
X(29)	29	-0.6000	0.4567	0.5624	0.0044	0.3344	-0.2551	0.1882	0.0109	0.4512	0.4351
X(30)	30	0.0615	0.4086	0.3470	0.4443	0.2972	0.4385	0.0674	0.4759	0.6299	0.6661
X(31)	31	-0.2267	0.7792	0.5821	0.1628	0.3469	0.3882	-0.0979	0.5499	0.8505	0.8292
X(32)	32	-0.2941	0.7029	0.6174	0.2224	0.3607	0.3252	-0.0819	0.4963	0.9075	0.8537
X(33)	33	-0.2812	0.5046	0.6989	0.1329	0.2597	0.1827	-0.0091	0.3447	0.8057	0.6568
X(34)	34	-0.3611	0.6632	0.6159	0.2382	0.3725	0.2290	-0.1665	0.3064	0.8697	0.7869
X(35)	35	-0.4181	0.2732	0.5075	0.2497	0.2033	-0.1400	-0.0720	-0.0644	0.4273	0.3672
X(36)	36	-0.4223	-0.1975	0.3933	-0.1832	-0.3100	-0.4680	0.1132	-0.4728	0.1741	0.0632
X(37)	37	-0.3320	0.3316	0.7327	0.1744	0.3251	-0.0345	-0.0487	0.1056	0.6790	0.5359
X(38)	38	-0.4186	0.1915	0.7184	0.0350	-0.0907	-0.3363	0.0664	-0.1265	0.6080	0.3977
X(39)	39	0.1667	0.3544	-0.2809	0.4628	0.4234	0.3232	-0.1578	0.1429	-0.0881	0.1253
X(40)	40	0.3591	0.1722	-0.3813	0.2266	0.4158	0.6589	-0.2845	0.4884	-0.0134	0.1679
X(41)	41	-0.0820	0.5061	-0.1514	0.1220	0.6398	0.3829	-0.1085	0.3881	-0.0431	0.1891
X(42)	42	-0.4300	0.6361	0.7532	0.2464	0.0935	0.0879	-0.0378	0.3391	0.9145	0.7457
X(43)	43	0.2856	-0.7203	-0.5946	-0.1920	-0.2437	-0.2398	0.0671	-0.4373	-0.8243	-0.7516
X(44)	44	-0.4242	0.3460	0.6212	-0.0614	-0.0050	-0.2587	0.1084	0.0253	0.7258	0.5169

Table 3.3 (part II):

		X(21) 21	X(22) 22	X(23) 23	X(24) 24	X(25) 25	X(26) 26	X(27) 27	X(28) 28	X(29) 29	X(30) 30
X(21)	21	1.0000									
X(22)	22	-0.2929	1.0000								
X(23)	23	-0.6497	0.1605	1.0000							
X(24)	24	0.8251	-0.4741	-0.4363	1.0000						
X(25)	25	0.7720	-0.5365	-0.5861	0.7092	1.0000					
X(26)	26	0.5763	-0.5351	-0.3334	0.6520	0.6012	1.0000				
X(27)	27	0.4793	-0.6165	-0.1357	0.6910	0.6163	0.5370	1.0000			
X(28)	28	-0.6248	0.1447	0.5045	-0.4710	-0.5954	-0.2625	-0.2152	1.0000		
X(29)	29	0.4751	0.0940	-0.4292	0.2871	0.2276	0.3511	-0.0758	-0.0223	1.0000	
X(30)	30	0.4678	-0.3818	-0.2520	0.3772	0.5642	0.5865	0.2948	-0.4006	0.2013	1.0000
X(31)	31	0.8951	-0.3931	-0.6495	0.7682	0.8210	0.6793	0.4546	-0.6795	0.3469	0.6477
X(32)	32	0.8387	-0.3284	-0.6131	0.7164	0.7266	0.7413	0.3894	-0.6463	0.4278	0.7174
X(33)	33	0.7468	-0.2048	-0.5039	0.6934	0.5762	0.5532	0.2639	-0.4785	0.3536	0.5975
X(34)	34	0.8119	-0.2343	-0.5447	0.6785	0.6560	0.6224	0.3204	-0.5917	0.4711	0.6065
X(35)	35	0.3132	-0.0369	-0.3954	0.1977	0.1774	0.2621	-0.2082	-0.1732	0.5702	0.2952
X(36)	36	-0.0328	0.4635	-0.0791	-0.2035	-0.1667	-0.1998	-0.5632	-0.1684	0.3491	0.1181
X(37)	37	0.5228	-0.1596	-0.3202	0.5244	0.3595	0.4565	0.1166	-0.2095	0.3778	0.4556
X(38)	38	0.2872	0.1414	-0.4054	0.1561	0.2348	0.2494	-0.1166	-0.3800	0.4084	0.4717
X(39)	39	0.0615	-0.5287	0.0033	0.1528	0.2629	0.2257	0.2969	0.0639	0.1021	0.1342
X(40)	40	0.0500	-0.6547	0.0270	0.2545	0.3000	0.3348	0.4321	-0.1429	-0.3714	0.2425
X(41)	41	0.2826	-0.4554	-0.3215	0.4273	0.3408	0.3932	0.4192	0.1541	0.2730	-0.0971
X(42)	42	0.7781	-0.1274	-0.6063	0.5908	0.6405	0.4861	0.3053	-0.5479	0.3846	0.5867
X(43)	43	-0.7954	0.2342	0.5679	-0.7137	-0.7009	-0.6120	-0.4786	0.5771	-0.3463	-0.5055
X(44)	44	0.5731	0.0931	-0.5395	0.3823	0.3166	0.2963	-0.0931	-0.4976	0.5344	0.3873

		X(31) 31	X(32) 32	X(33) 33	X(34) 34	X(35) 35	X(36) 36	X(37) 37	X(38) 38	X(39) 39	X(40) 40
X(31)	31	1.0000									
X(32)	32	0.9107	1.0000								
X(33)	33	0.7443	0.7587	1.0000							
X(34)	34	0.8106	0.9324	0.7150	1.0000						
X(35)	35	0.3116	0.4527	0.4007	0.5572	1.0000					
X(36)	36	-0.0059	0.1479	0.0762	0.2195	0.4242	1.0000				
X(37)	37	0.5338	0.6183	0.7423	0.6286	0.3269	0.1680	1.0000			
X(38)	38	0.4088	0.5514	0.4793	0.5584	0.4266	0.5657	0.4933	1.0000		
X(39)	39	0.0791	-0.0104	-0.1315	0.0538	0.0291	-0.2919	-0.1284	-0.2508	1.0000	
X(40)	40	0.1208	0.1448	0.0130	0.1254	-0.1668	-0.4388	-0.0429	-0.2829	0.3097	1.0000
X(41)	41	0.2189	0.1106	0.0134	0.0780	0.1128	-0.4827	-0.1418	-0.4377	0.5403	0.3213
X(42)	42	0.7995	0.8024	0.7930	0.7764	0.3599	0.1407	0.6327	0.6394	-0.1343	-0.0953
X(43)	43	-0.8188	-0.8300	-0.6684	-0.7630	-0.2267	0.1336	-0.5177	-0.4360	-0.0190	0.0023
X(44)	44	0.5797	0.6546	0.6404	0.6040	0.4638	0.5261	0.6026	0.7709	-0.3766	-0.3889

		X(41) 41	X(42) 42	X(43) 43	X(44) 44
X(41)	41	1.0000			
X(42)	42	-0.0807	1.0000		
X(43)	43	-0.1648	-0.8136	1.0000	
X(44)	44	-0.3024	0.6773	-0.5091	1.0000

Table 3.3 (part III):

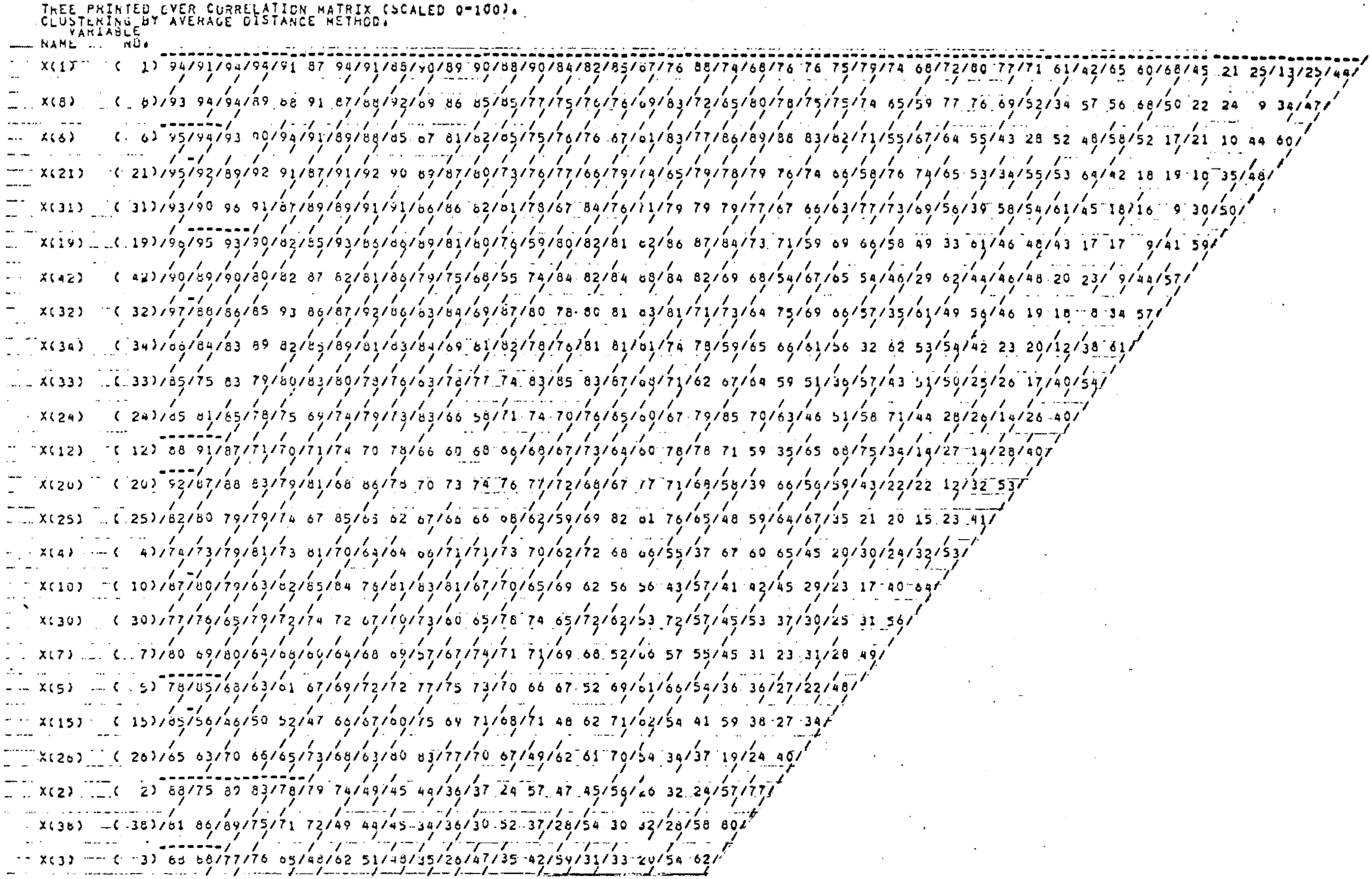


Figure 3.1 The clustering solution for the E3 results in the form of a tree (continued on the next page).

```
X(13)  ( 13) 92/87/79 75/43 51 49 37/31 26/47/36/42/56 26 31 20 56/70/
X(44)  ( 44)/81/77/75/45 52/46 38/31 29 47 32/35/56 24/26/25 55 17/
X(37)  ( 37)/69 66/56/55/56 48 48 33/59/44/43/48/34/40/24 42 56/
X(29)  ( 29) 79/47/51 47/38 31 20 51/55/64/60 29/49/33/55/67/
X(35)  ( 35)/45 46 40 43 41/29/62 51/56/46 30 41/39 48/71/
X(9)   (  9) 77 80/74/76/78/74/73/63/54/54 47 40 18/30/
X(18)  ( 18) 68/86 74/62/60/57/69/43/41 36/26/18 26/
X(27)  ( 27)/80/72/66/60 65/71/47 44 39/26 20 22/
X(16)  ( 16)/83/69/67/66 69/44 47/36/38/20/27/
X(40)  ( 40)/68/61/65 66/36/51 43/50/17/28/
X(11)  ( 11)/57 58 46/52 76/53/64/35/29/
X(14)  ( 14) 73 56/51 44 53/40 26/41/
X(39)  ( 39) 77/42/50/53/49/24 35/
X(41)  ( 41)/45 34/58 42/27 26/
X(17)  ( 17) 61 61 53/59 56/
X(23)  ( 23) 75 78/59 46/
X(28)  ( 28) 79/57/42/
X(43)  ( 43)/62 57/
X(22)  ( 22) 73/
X(36)  ( 36)/
```

AN EXPLANATION OF THE VARIABLE CLUSTERING PROCESS SHOWN IN THE TREE PRINTED ABOVE

THE PROCESS BEGINS WITH THE CLUSTER CONSISTING OF VARIABLE X(21) (21), THE 4TH VARIABLE LISTED IN THE TREE.

THIS CLUSTER JOINS WITH THE CLUSTER BELOW IT CONSISTING OF THE VARIABLE X(31) (31).
THE NEW CLUSTER IS INDICATED ON THE TREE BY THE INTERSECTION OF THE DASHES BEGINNING ABOVE VARIABLE X(21) (21)
WITH THE SLASHES STARTING NEXT TO VARIABLE X(31) (31).

THIS CLUSTER JOINS WITH THE CLUSTER ABOVE IT CONSISTING OF THE VARIABLE X(6) (6).
THE NEW CLUSTER IS INDICATED ON THE TREE BY THE INTERSECTION OF THE DASHES BEGINNING ABOVE VARIABLE X(6) (6)
WITH THE SLASHES STARTING NEXT TO VARIABLE X(31) (31).

THIS CLUSTER JOINS WITH THE CLUSTER ABOVE IT CONSISTING OF THE VARIABLES X(1) (1) DOWN TO X(8) (8).
THE NEW CLUSTER IS INDICATED ON THE TREE BY THE INTERSECTION OF THE DASHES BEGINNING ABOVE VARIABLE X(1) (1)
WITH THE SLASHES STARTING NEXT TO VARIABLE X(31) (31).

THE PROCESS CONTINUES UNTIL EACH VARIABLE IS JOINED TO AT LEAST ONE OTHER VARIABLE

Figure 3.1 (Part II).

Table 3.5 gives the correlations between these six scales and the 27 physical features derived in chapter two.

Features	Preferences 1	2	3	4	5	6
1	.882	.295	-.769	-.005	-.012	.459
2	.559	-.012	-.550	-.060	.377	.549
3	.072	.382	.024	-.354	.360	-.237
4	.653	.452	-.646	-.231	.291	.176
5	.492	-.151	-.446	.085	-.041	.463
6	-.904	-.374	.828	.128	-.010	-.291
7	.137	.043	.109	.070	-.239	.031
8	.031	.206	.178	-.072	-.138	-.051
9	.338	-.078	-.459	.251	-.330	-.064
10	.528	.640	-.446	-.282	.185	-.019
11	-.103	-.176	-.012	.324	-.301	-.343
12	-.514	.023	.437	.018	-.308	-.550
13	.249	-.052	-.188	.210	-.526	-.004
14	.080	-.158	-.113	-.024	.487	.371
15	-.260	-.254	.180	-.038	.221	.062
16	-.078	.187	.052	-.130	.323	-.044
17	.128	-.395	-.003	.610	-.512	.346
18	-.026	.082	-.109	-.277	.039	-.222
19	-.035	-.289	-.120	.124	-.213	.083
20	-.093	-.361	.211	.352	-.092	.218
21	.314	-.003	-.257	.173	-.391	.065
22	-.143	.025	.074	-.194	.254	-.073
23	-.169	-.128	.194	-.095	.363	.072
24	.541	.119	-.443	.202	-.457	.090
25	.763	.348	-.722	-.201	.423	.446
26	-.227	.284	.271	-.085	.004	-.329
27	.878	-.002	-.693	.269	-.193	.635

Table 3.5 Correlations between the six preference scales and the first 27 physical features of the Walsh stimuli.

The existence of the six clearly defined clusters of preference response indicates that individual differences are an important factor in preference judgements involving the Walsh stimuli. Further studies are now needed, using similar data analytic methods to those used for E3 above, to replicate this finding. It would be particularly interesting to see whether similar results are obtained using other stimulus sets, and, if so, whether there are consistencies in the type of preference scale that a given individual generates across different stimulus sets. Once the necessary studies have been done it should then be possible to determine the attributes of the individual which are predictive of the preference that (s)he has.

Conceptual Ranking and Preference

Figure 3.2 indicates that there is a certain amount of error in the conceptual ranking method used in E3 which is marked at moderate levels of preference.

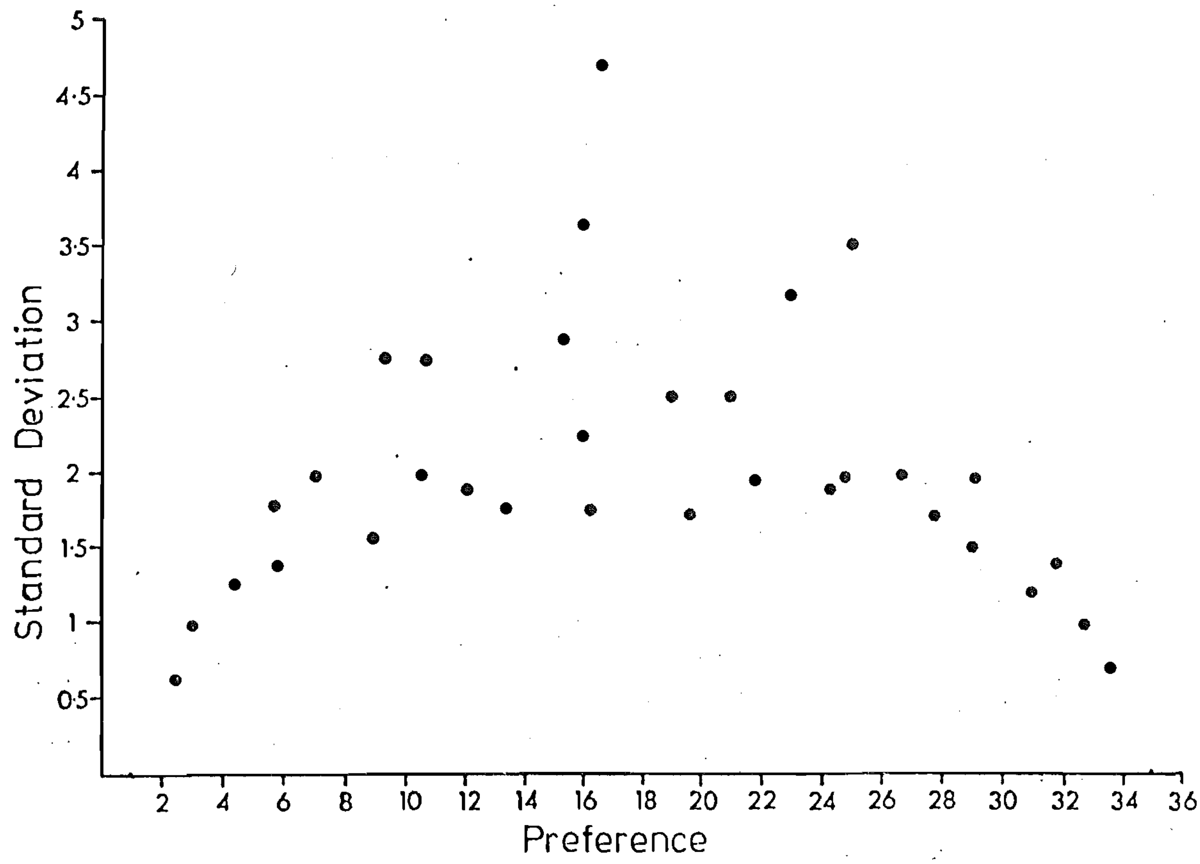


Figure 3.2 A scatterplot of the implied preference rank in a 7 x 5 conceptual grid (x-axis) against the standard deviation (y-axis) of that rank as estimated by Monte Carlo simulation (100 trials).

One method of reducing the effect of this error is to average over the participants. Thus the values in preference scale one should fairly closely reflect the participants' judgements (averaged over 27 cases), while the other preference scales will probably contain some error due to the conceptual ranking task. The advantage of using conceptual ranking followed by the cluster analysis of the results is that this procedure should avoid apparently paradoxical results such as those obtained by Gregson (1968). Gregson showed that his results could be explained by a judgement model in terms of shifting frames of reference for individual subjects. In the present study conceptual ranking constrains the individual to respond within a consistent frame of reference, while cluster analysis partitions the individuals into homogeneous subgroups. Thus, the present procedure should alleviate some of the data collection problems which have previously had to be dealt with in the analysis of results. In particular, it should remove the problem of aesthetic fatigue in ranking (West and Bendig, 1954) since all the stimuli are effectively ranked simultaneously, rather than successively over time.

Preference and Complexity

A relationship between complexity and preference has been suspected for a number of years, with a variety of models being suggested (Smets, 1973). Some of the major models are summarised in Figure 3.3 (after Smets, 1973; Figure 2.6).

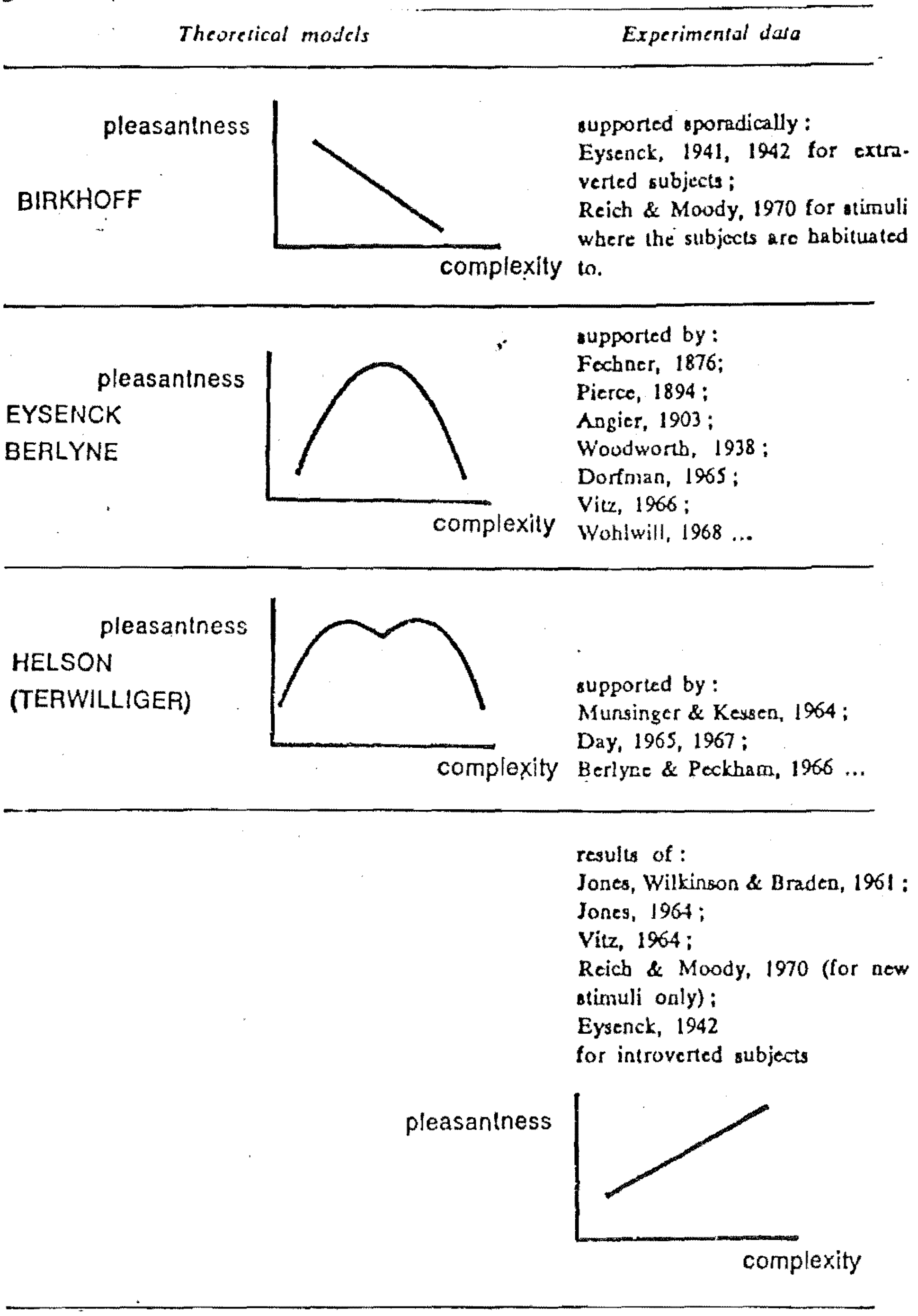


Figure 3.3 A schematic representation of four models of the relationship between stimulus complexity and aesthetic preferences, each accompanied by the studies where that relationship was found (after Smets, 1973; Figure 2.6).

It can be seen that a model such as that of Helson (1964) where preference is determined by distance from an ideal point can explain the positive and negative relationships shown at the bottom and top of Figure 3.3, with the ideal point being assumed to be outside the range of stimuli used. The advantage of such a model is that it can be adapted to explain changes in preference structure during learning:

"According to a thory proposed by Dember and Earl, the effective complexity of a stimulus is derived from its perceived characteristics in interaction with the perceiver's own ability to appreciate the potential complexity of the stimulus. In addition, the theory states that each individual has a preferred, or ideal, level of complexity on each attribute. The individual is most responsive to stimuli at this level. Given the opportunity, the individual will also have perceptual contact with stimuli below and above the ideal level, including stimuli that are slightly above his own momentary level. These latter stimuli, called *pacers,* have the effect of pulling the individual's ideal up to their level. Thus, under the proper circumstances, an individual's ideal is continually increasing, perhaps to a level which is limited by his hereditary endowment. As the individual's ideal increases, of course, his attention will be directed primarily toward stimuli of increasing complexity."

— Dember,(1960, p.374).

This notion of preference being determined by distance from an ideal point has been formalised in terms of unfolding theory (Coombs, 1964) and the theory of single-peaked functions (Coombs & Avrunin, 1977). The unfolding theory approach will be looked at more closely in a later section.

Dorfman and McKenna (1966) tested the hypothesis that level of preference for patterns is a function of uncertainty defined in terms of matrix grain. They obtained their preference data using paired comparison judgements between 36 stimuli which were essentially black and white checkerboard patterns. In analysing the results of E3 there was consequently interest in whether or not results would replicate the findings of Dorfman and McKenna. Such replication would suggest that the Walsh stimuli (at least in respect to the matrix grain-preference relationship) are perceived as a subset of the general set of black and white checkerboard patterns, rather than being an atypical stimulus set which was perceived in an unusual fashion.

Dorfman and McKenna divided their participants into six groups according to the participants' most preferred level of uncertainty among the six levels of stimulus uncertainty which were used in the experiment. They found differences between the six groups as shown in Figure 3.4.

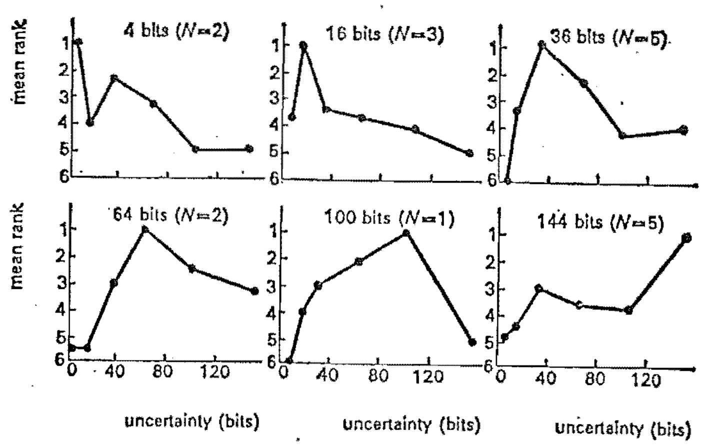


Figure 3.4 a. The relationship (for art students) between level of uncertainty and mean preference ranking for six groups, ordered according to most preferred level of uncertainty (after Dorfman and McKenna, 1966, Figure 5).

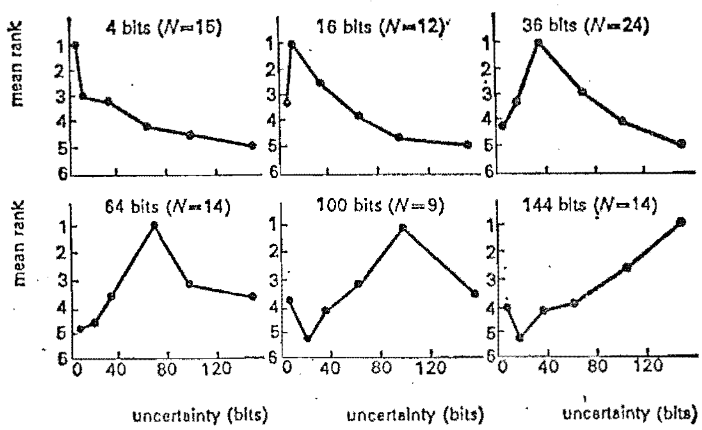


Figure 3.4b The same relationship as in figure 3.4a except that non-art students were used (after Dorfman and McKenna, 1966, Figure 3).

Dorfman and McKenna used an informational measure of uncertainty, that is:

$$H = -\sum_i \sum_j pij \, \mathrm{Log}_2 \, pij \qquad (3.1)$$

Where pij is the probability that the ijth cell of the matrix will be filled in.
Their six levels of information correspond to 4, 16, 36, 64, 100, and 144 bits of information respectively. The derivation of an equivalent informational measure for the Walsh stimuli is not obvious, because of their construction from Walsh functional, rather than probabilistic considerations. One method is to use the product of the row and column sequencies for each stimulus as an estimate of informational uncertainty. This method generates a further feature for the Walsh stimuli which is shown in Table 3.6.

Stimulus Number	Two dimensional Sequency
2	2
3	3
4	4
5	5
6	6
7	7
8	8
10	4
11	6
12	8
13	10
14	12
15	14
16	16
19	9
20	12
21	15
22	18
23	21
24	24
28	16
29	20
30	24
31	28
32	32
37	35
38	30
39	35
40	40
46	36
47	42
48	48
55	49
56	56
64	64

Table 3.6 Two dimensional (2-D) sequency values for the 35 Walsh stimuli.

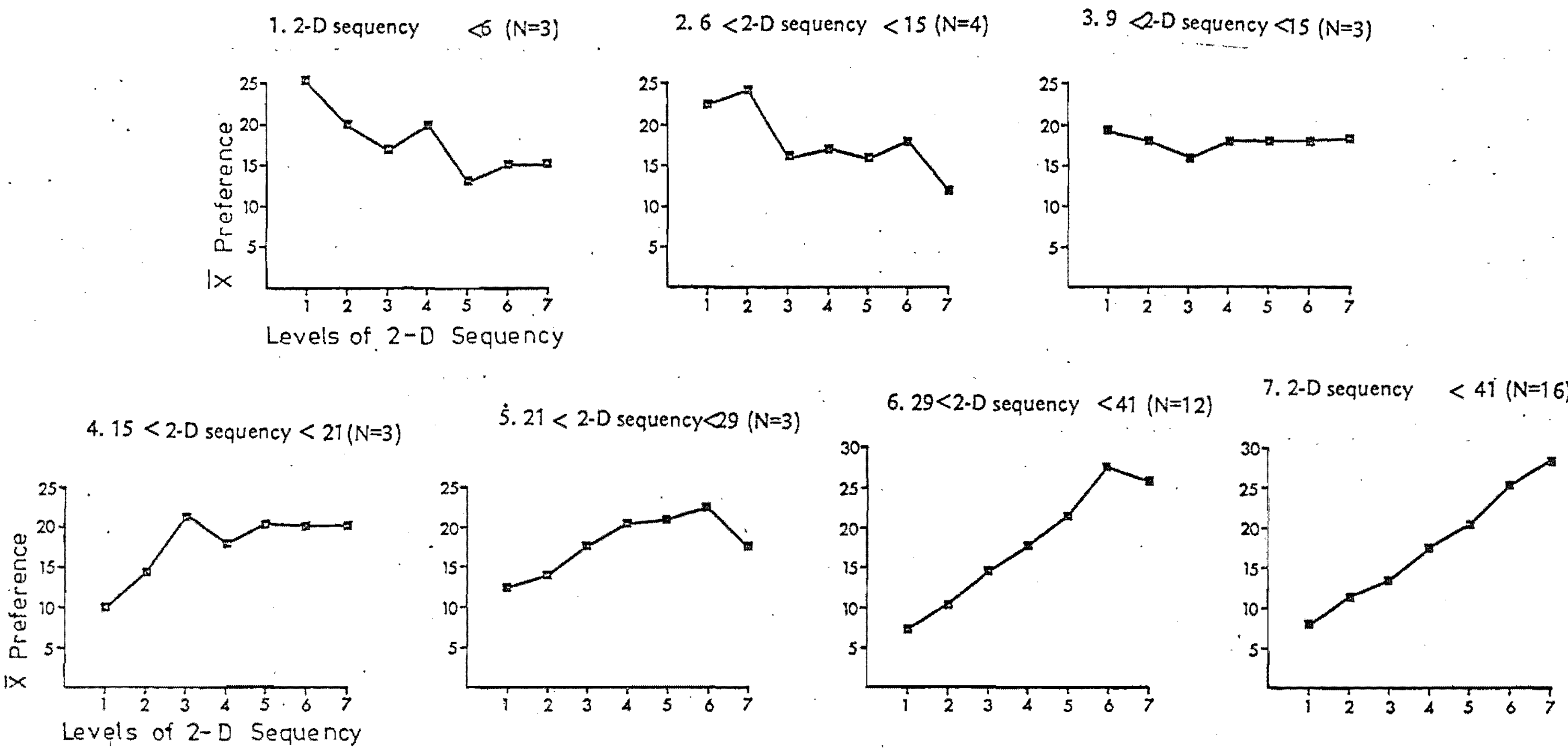


Figure 3.5 The mean preference ranks across seven levels of 2-D sequency for seven sets of participants grouped according to their most preferred level of 2-D sequency.

This feature(which shall be referred to as two-dimensional sequency) was then divided into seven levels with five of the 35 Walsh stimuli in each level. The 44 participants in Experiment E3 were then classified according to which of the seven levels of two-dimensional sequency their most preferred stimulus belonged in. The mean rank of preference, for each of the seven classifications, as a function of two-dimensional sequency is plotted in Figure 3.5. This figure can be directly compared with the results of Dorfman and McKenna given in Figure 3.4. In general, Figure 3.5 is consistent with the theory of Dember and Earl (1957) in that preference decreases with distance from the ideal point (where distance is along the dimension of uncertainty, or, in the present case, two-dimensional sequency). The major discrepancy involves the three participants who had a most preferred stimulus with a two-dimensional sequency of between nine and fourteen. Two of these individuals were from cluster two and one from cluster four. This indicates the possibility that there is a group of participants who do not use uncertainty in making judgements of preference between the Walsh stimuli. This possibility will be pursued further in the next section.

Individual Differences, Preference, and Complexity

Most, if not all, previous studies investigating the relationship between complexity and preference have defined complexity in terms of certain physical features of the stimuli used. Experiment E2 (outlined in Chapter Two) derived an empirically-based measure of complexity for the Walsh stimuli. Figure 3.6 shows the relationship between subjective complexity and preference for the first three preference scales. It can be seen that the curves of the final two plots in Figure 3.5are similar to the complexity-preference relationship for preference scale one.

Preference scale three, on the other hand, would be predicted by Dember and Earl's (1957) theory (if the perceived complexity is regarded as being closely related to stimulus uncertainty) where the ideal point corresponds to a stimulus with low uncertainty/complexity.

Preference scale two does not appear to have any systematic relationship with complexity. The relationship between preference, complexity, and a number of physical features of the Walsh stimuli for each of the six preference scales (as derived previously from the cluster analysis) is studied further in the next section.

Stepwise Regression Analysis

The previous discussion considered the relationship between complexity and preference. It is also possible that other physical features may have an effect on preference. Stepwise multiple regression (with forward inclusion, BMD02R; Dixon, 1973) was used to fit models relating each of the six preference scales to the physical features of the Walsh stimuli. The Stepwise regression for preference scale one produced the following predictor equation which took up 90% of the variance in the preference scale values.

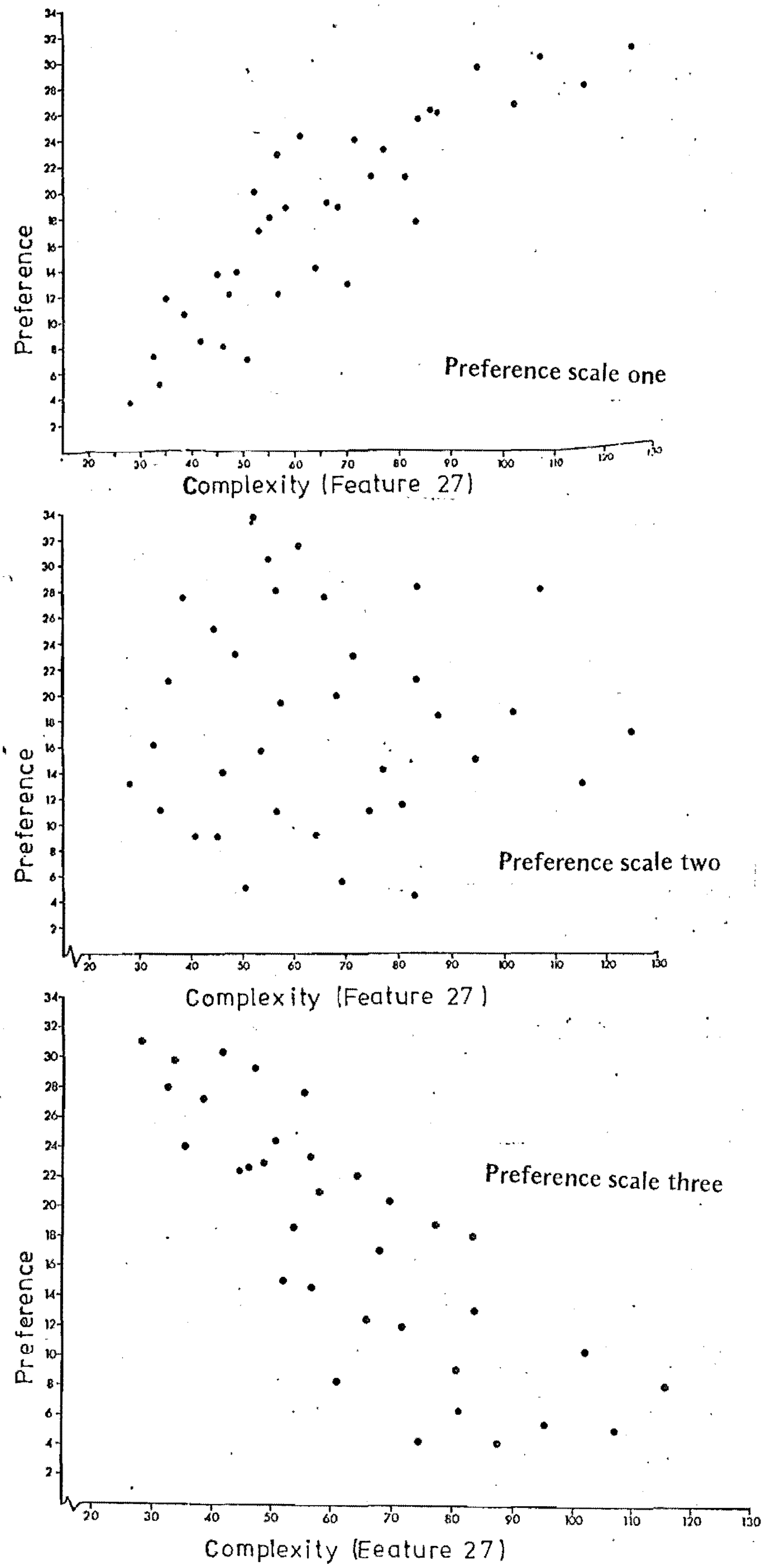


Figure 3.6 Scatterplots of the relationships between complexity and the rankings on preference scales one, two, and three (the multiplicative scale of complexity derived in Chapter Two is used here).

Pref 1 = .14 complexity – 7.48 graininess + 16.3 (Equation 3.2).

This equation is difficult to interpret (c.f. Draper and Smith, 1966) because of the high intercorrelations between the two independent variables. (The correlation between complexity and graininess is –.77).

A second Stepwise regression was run for preference scale one, with graininess (Feature 6) being removed from the set of independent variables. This produced the following fitted equation which accounted for 87% of the criterion variance:

Pref 1 = 2.77 sequency + 1.41 component 2 + 9.14 (Equation 3.3).

Equation (3.2) uses complexity and graininess (another measure of complexity; c.f. Dorfman and McKenna, 1966) to predict preference, while Equation (3.3) uses two of the physical features which do not have an explicit relationship with complexity, (although they will have some relationship defined statistically over the finite set of stimuli used). Consequently, it can be seen that while preference is explainable in terms of complexity, it is also explained adequately by two other physical features of the Walsh stimuli.

The Stepwise regression for preference scale three produced the following equation which explained 77% of the criterion variance:

Pref 3 = 12.78 graininess + 1.21 grain variance + 3.04 (Equation 3.4).

A second analysis, with graininess removed, produced the equation:

Pref 3 = –2.38 sequency – 1.66 component 2 + 25.61 (Equation 3.5).

which took up 71% of the variance.

Stepwise regression analyses[1] were also run for the other four preference scales but graininess was not an important predictor for these scales. 64% of the variance in preference scale six was taken up by column sequency, component one, and the number of squares in the stimulus, while scales two, four and five had only 58%, 46% and 43% respectively of their variance accounted for by the final regression equation. In view of this low predictability of scales two, four, and five in terms of the physical features, and the relatively high predictability of scales one, three, and six, the next section will look at an alternative method of characterising scales two, four, and five. The remainder of this section will discuss the implications of the Stepwise Regression results.

1. In all the Stepwise regression analyses used with the preference scales, F to include was set at 5.0 and F to remove was set at 3.0. These parameter settings meant that the variable would be included only if the increase in variance accounted for by it was significant at approximately the .01 level.

The Stepwise regression analyses of the six preference-scales show that some, but not all, types of preference can be related statistically to complexity. The general substitutability of complexity with other features in the predictor equation (notably column sequency and component 2) suggests that preference judgements may be made directly on the physical features, rather than being mediated by complexity. The close relationship of complexity to other physical features in checkerboard stimuli makes such an hypothesis difficult to test, however. It also seems that if the various measures are highly intercorrelated then a participant who changes the physical bases of his preference from one stimulus pair to another will superficially appear to be using a constant basis, even though in an imprecise fashion (Gregson, 1963). If and when such behavious occurs, it will be possible to identify neither the stationarity of response strategies nor the features which mediate the preference judgements. It thus appears that the fitting of regression models can only provide approximate models of preference judgements which cannot necessarily identify non-stationarity in responding.

Preference and Cognitive Structure

The preceding Stepwise regression analysis could characterise adequately only three of the six preference scales. The notion of distance from an ideal point will now be used to describe how the other three preference scales may be derived using a particular theory of cognitive structure.

As Coombs (1964) has pointed out, theories of categorising based on an ideal viewpoint have been around at least since the time of Plato and have been espoused in slightly different forms by a number of psychological theorists (Bartlett, 1932; Hayek, 1952; Vernon, 1955; Bruner, Goodnow and Austin, 1956; and Miller, Galanter and Pribam, 1960). Rosch (1977) has recently outlined a version of the theory which has received a large amount of experimental support (e.g. Rosch and Mervis, 1975, and Rosch, Mervis, Gray, Johnson and Boyes-Braem, 1976). One of the major principles of Rosch's theory (Bruner, 1957 uses a similar argument) is the precept of cognitive economy, that is, the person seeks to abstract information from the environment as much as possible, without blurring over essential differences.

"On the one hand, it would appear to the organism's advantage to have as many properties as possible predictable from knowing any one property, a principle that would lead to formation of large numbers of categories with as fine discriminations between categories as possible. On the other hand, one purpose of categorisation is to reduce the infinite differences among stimuli to behaviourally and cognitively usable proportions."

— Rosch (1978, pp 28-29).

A basic notion that follows from the precept of cognitive economy is that two stimuli which are in the same category should be pairwise more similar than two stimuli which belong to two different categories. Reed (1973, Chapter 9) has outlined four models based on this consideration. These models represent stimuli in a multidimensional space. A prototype may then be defined as the best representative of a category and corresponds to Coombs' (1964) notion of an ideal point. Coombs interpreted the process of perception as involving relations between pairs of (multidimensional) points from distinct sets. On the one hand there are ideal points (prototypes) for the equivalence classes (categories) in the individual's cognitive structure, and on the other hand there are points for the stimulus inputs. Coombs (1964) described a perceptually-based theory of categorisation in the following manner:

"I have been describing the process of perception as if the individual searched his cognitive structure of ideal points, comparing the distance of a stimulus point vis-a-vis each such point until he found one *which satisfied the criterion for matching* —and thereby the stimulus is identified, labelled, and perceived. We might make this process more reasonable by adding that the stimulus somehow controls the space of relevant dimensions in which this search takes place."

— Coombs, (1964, pp.332-333).

Whitfield and Slatter (1979) suggest that aesthetic choice may reflect categorisation and prototypicality. However, if they are correct then it appears that the individual differences found in the pattern of preferences for the Walsh stimuli should reflect individual differences in category structure. Sorting tasks are one way of estimating category structure and these may be used to check for consistencies between category structures and preferences. A particular model of the relationship between cognitive structure and preference will now be suggested as an explanation for the implied preference scales two, four, and five.

Unfolding theory and preference scales two, four, and five.

Unfolding theory has been characterised as follows:

"The basic assumptions of the theory of preferential choice on which the unfolding technique in one dimension is based are as follows. Each individual and each stimulus may be represented by a point on a common dimension called a *J Scale* and each individual's preference ordering of the stimuli from most to least preferred corresponds to the rank order of the absolute distances of the stimulus points from the ideal point, the nearest being the most preferred. The individual's preference ordering is called an *I scale* and may be thought of as the J scale folded at the ideal point with only the rank order of the stimuli given in order of increasing distance from the ideal point. The data consists of a set of I scales from a number of individuals, and the analytical problem is how to unfold these I scales to recover the J scale."

— Coombs, (1964, p.80).

Table 3.7 shows the intercorrelations between the six preference scales.

Preference Scales	Features 1	3	6	2	4	5
1	-					
3	-.817	-				
6	.542	-.322	-			
2	.371	-.257	-.224	-		
4	-.126	.099	.290	-.792	-	
5	.107	-.064	.138	.463	-.586	-

Table 3.7 Matrix of intercorrelations between the six preference scales where the scales have been re-ordered to emphasise clustering.

It can be seen that the pattern of intercorrelations between the preference scales indicates the existence of two clusters, the first consisting of scales one, three and six, and the second consisting of scales two, four and five. The two clusters distinguish those scales which could be adequately accounted for by the Stepwise regression analysis (one, three, and six) and those which could not (two, four and five). The first cluster of scales can be viewed as having a J scale which corresponds to complexity, or one of the closely related physical features (such as column sequency and average grain). The high negative correlation between scales one and three indicates a reverse ordering (approximately) which will arise when the ideal points are located at opposite ends of the J scales (Coombs, 1964). Figure 3.7 shows the effect of the ideal points being located at either end of the complexity (J) scale for preference scales one and three. The positive correlation between scales six and one and the negative correlation between scales six and three implies the following placement of the ideal points on the J scale.

I_3 I_6 I_1

2 10 4 3 12 8 28 11 5 16 19 64 7 22 6 13 20 24 29 15 21 23 37 40 14 56 31 30 22 48 46 39 38 55 47

Figure 3.7 A schematic representation of the ideal points for preference scales one, three and six located on the scale of complexity.

The relatively large intercorrelations between preference scales two, four, and five suggest that they may also be located on a common J scale. Table 3.8 gives the rankings of each of the 35 Walsh stimuli on preference scales two, four, and five.

Preference Rank	Preference Scales 2	4	5
1	28	14	8
2	64	15	32
3	19	22	64
4	32	47	7
5	37	11	5
6	55	30	28
7	10	46	16
8	29	6	56
9	12	38	6
10	40	31	40
11	16	20	39
12	8	13	10
13	56	3	37
14	21	5	55
15	24	2	48
16	38	55	4
17	48	7	19
18	46	23	47
19	4	21	24
20	20	48	23
21	39	56	21
22	23	24	12
23	11	10	11
24	47	8	20
25	2	32	29
26	30	4	46
27	31	39	31
28	13	19	15
29	3	28	22
30	15	16	2
31	7	40	3
32	5	29	13
33	14	12	38
34	6	37	30
35	22	64	14

Table 3.8 Rankings of 35 Walsh stimuli on preference scales two, four, and five.

According to unfolding theory, the I scale must end in either one of two different stimuli, those on the ends of the J scale (Coombs, 1964, p.86). Allowing for some noise in the data, Table 3.8 indicates that stimulus #14 is at one end of the J scale (and corresponds to the ideal point for scale four) while stimulus #64 is at the other end of the scale. Preliminary analysis indicates that the unfolding model appears to account for preference scales two, four and five quite well, although no attempt will be made to derive the underlying J scale (the algorithm is given in Coombs, 1964) in this thesis. A rough estimate of the J scale, with the ideal points for preference scales two, four and five, is given in figure 3.8.

Figure 3.9 (next page) is a visual representation of preference scale two. This appears to be the J scale underlying preference scales two, four and five.

		I_4																						I_5				I_2						
22	6	14	15	30	47	11	31	13	20	2	7	5	46	38	3	23	48	56	39	4	24	21	55	8	10	32	19	28	37	29	64	12	16	40

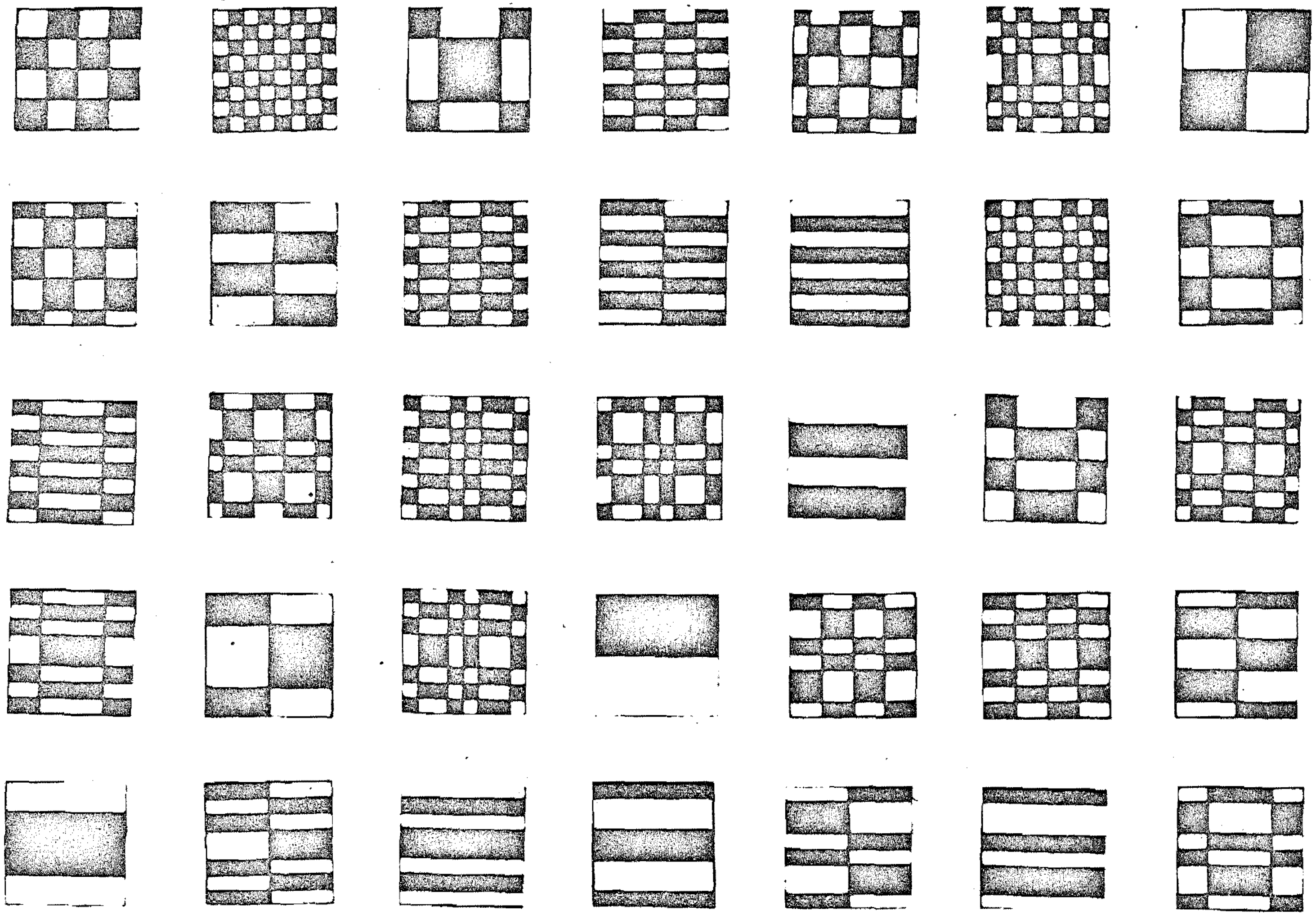
Figure 3.8 The approximate J scale and ideal points for preference scales two, four, and five.

Summary

The present chapter has shown consistent individual differences in preferences for the Walsh stimuli. The reasonable fit[1] of both unfolding theory (for all the scales) and regression models (for at least three of the scales) appear to validate the conceptual ranking procedure as a heuristic method of unidimensional scaling. The features derived in chapters two and three may now be used to investigate quantitatively the effects of task demands on perceptual and cognitive judgement. Such an investigation will be carried out in the following three chapters.

1. The six homogeneous groups of participants and the interpretable nature of their preference scales are the best evidence for the validity of the preference scales derived in E3.

Figure 3.9 A visual representation of preference scale two with order of preference decreasing across the rows. The top row shows the seven most preferred stimuli, the second row gives the next seven and so on.



CHAPTER IV

INTRODUCTION

Chapter One outlined some of the elements in a theory of learning. This theory will be used as a framework within which the results of subsequent experimentation may be interpreted. The present chapter will examine the role of attentional strategy in perceptual-cognitive processing. The term perceptual-cognitive will be used to describe the processing required of the participant during the experiments used in this thesis, as the experimental tasks generally require both perceptual and cognitive processing. In addition, there appears to be no point in making a clear (and somewhat arbitrary) distinction between the two types of processing in this thesis.

Norman's Model of Perceptual-Cognitive Processing

Waugh and Norman (1965) evaluated the notion of primary and secondary memory. Norman (1968) implied that the two storage systems might be logically, but not physically different.[1] He suggested a dual-process storage system where secondary and primary storage are different properties of the same system. The different properties of the two storage models arise because primary traces are continually changing whereas secondary traces are passive and permanent. Norman described the relationship between primary and secondary storage in the following manner:

"There are two forms of storage: primary and secondary. The two forms are different aspects of one large storage system. Primary storage is the temporary activation of this large storage by the sensory inputs. Secondar y storage is a long term excitation."

— Norman, (1968, p.535).

In terms of the theory outlined in chapter one, the large storage system consists of stored representations while the primary storage consists of the activated memory schemata. Figure 4.1 shows Norman's (1968, figure 1) model of selection and attention.

1. Norman (1968) appears to be making a hardware-software distinction but the point will not be taken up here. The power of the computer metaphor (see Waterman and Hayes-Roth, 1978, for current examples of computer-based cognitive modelling) needs to be balanced by empirically-based analyses of perception and cognition (the approach taken in this thesis).

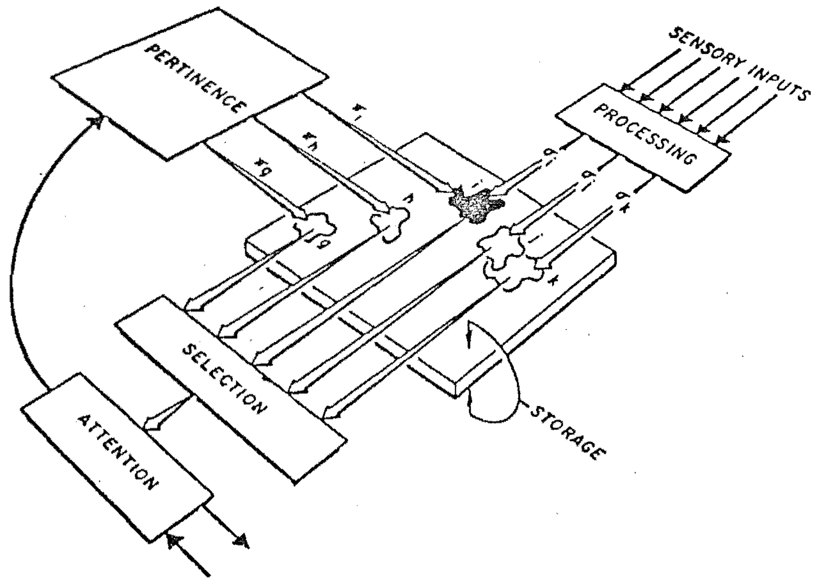


Figure 4.1 An outline of selection and attention.. (Sensory inputs, σ , after undergoing physiological processing, excite their representations in storage – i, j, and k in the figure. Simultaneously, higher-level cognitive factors have determined what stored representatives are most pertinent to the psychological processes that are going on at the moment. These representatives – g, h, and i in the figure – receive a pertinence input, π . The item which is selected for further processing by an attention mechanism corresponds to the stored representation which received the greatest combination of pertinence and sensory activation– the shaded item, i, in the figure). "After Norman (1968, figure 1).

Attention and Pertinence form two additional components of Norman's (1968) model. The caption of Figure 4.2 shows that Norman's attention mechanism operates *after* the sensory input has been recognised (has made contact with long term storage). This point that attention occurs *after* recognition of the stimulus has also been made by Keele:

"The retrieval of information stored in memory and triggered by an external stimulus does not require attention. Subsequent operations, in contrast, are attention-demanding. To say that memory retrieval does not require attention is simply to say that if more than one source of sensory information impinges on a person at the same time, all the sources activate information stored in memory without interfering with each other."

– Keele (1973, pp 137–138).

The approach to be adopted here is that selective attention is a 'control process' rather than a result of structural characteristics of the processing system:[1]

1. See Broadbent (1971) for the alternative point of view.

" . . . Selective attention is largely defined in terms of control processes. Selectivity due to structural characteristics of the processing system, even structure that has been learned, is not considered to be an attentional process . . . (A)ttentional selectivity is largely the result of accentuation of certain informational elements through the use of control processing . . ."

—Shiffrin & Schneider,(1977, pp177—178).

The distinction between controlled and automatic processing has been described by Shiffrin and Schneider (1977, pp 159—160). Figure 4.2 (after Shiffrin and Schneider, 1977, figure 11) gives a schematic representation of a model for automatic and controlled processing.

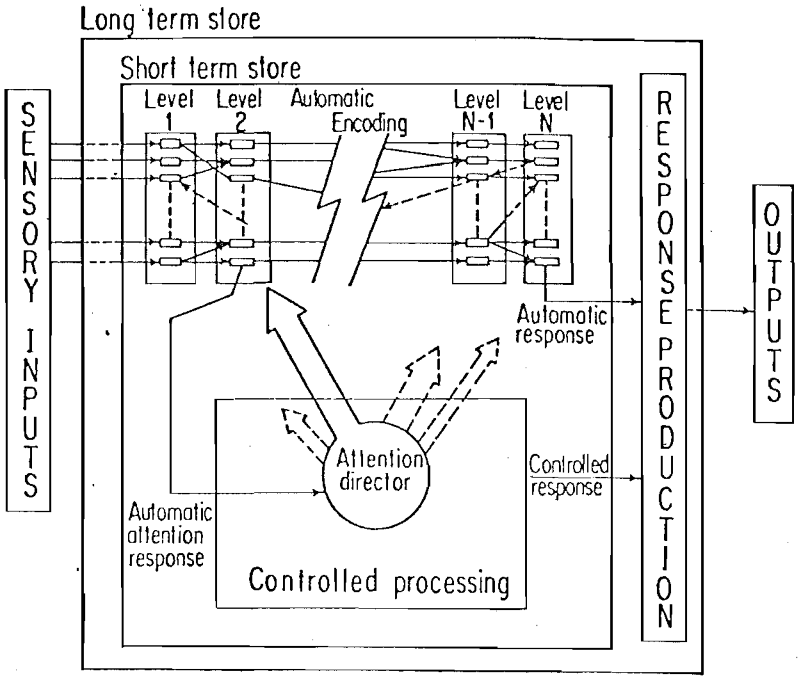


Figure 4.2 A model for automatic and controlled processing during tasks requiring detection of certain input stimuli. Short-term store is the activated subset of long-term store. N levels of automatic encoding are shown, the activated nodes being depicted within each level. The dashed arrows going from higher to lower levels indicate the possibility that higher level features can sometimes influence the automatic processing of lower level features. The solid arrow from a node in Level 2 to the attention system indicates that this node has produced an automatic-attention response, and the large arrow from the attention system to Level 2 indicates that the attention system has responded. The arrow from level N to the Response Production indicates that this node has called for an automatic overt response, which will shortly be executed. The arrow from Controlled Processing to the Response Production indicates the normal mode of responding in which the response is based on controlled comparisons and decisions. Were it not for the automatic responses indicated, detection would have proceeded in a

serial, controlled search of nodes and levels in an order chosen by the subject. After Shiffrin and Schneider (1977, figure 11).

Figure 4.2 explains the activation of memory schemata (short term store) within the system of stored representations (long term store) in terms of the attention directors. Shiffrin and Schneider, however, do not appear to have considered the notion of representational structure as characterised in Chapter one. The following argument suggests that some concept such as that of representational structure will be necessary to account for a participant's memory of previous trials during an experiment. Consider the state of a person's memory during the course of a series of pairwise similarities judgements. (S)he will have a long term store which consists of information accumulated prior to the experiment, as well as the activated memory schemata which result from the effect of sensory input in the presence of the memory context (Morton, 1969). In addition, (s)he will have some kind of representation of the stimuli which has been built up during the course of the experiment, and this will be more permanent than the activated memory schemata which necessarily change from trial to trial. This additional notion of representational structure appears metatheoretically indispensable in view of the following (rather transient) definition of short term store:

"Short term store is the labile form of the memory system and consists of the set of concurrently activated nodes in memory."

— Shiffrin and Schneider (1977, p.157).

Response Derivation and Logogens

Shiffrin and Schneider (1977) recognised the existence of response production, without attempting to give a description of what it involves. It was suggested in chapter one of this thesis that the overt responses made by a participant during an experiment are derived by operations upon the representational structure in order to make the responses which are appropriate within the experiment. Morton's (1969) Logogen model is one theory which explicitly considers response derivation. The basic idea of the Logogen model is as follows:

"The Logogen is a device which accepts information from the sensory analysis mechanisms concerning the properties of linguistic stimuli and from context-producing mechanisms. When the logogen has accumulated more than a certain amount of informa tion, a response (in the present case the response of a single word is made available."

—Morton (1969, p.176).

An appropriately modified form of the Logogen model appears to be one way of characterising the response derivation process. The specification and testing of such a model is likely to be a difficult task and will not be considered further in this thesis.

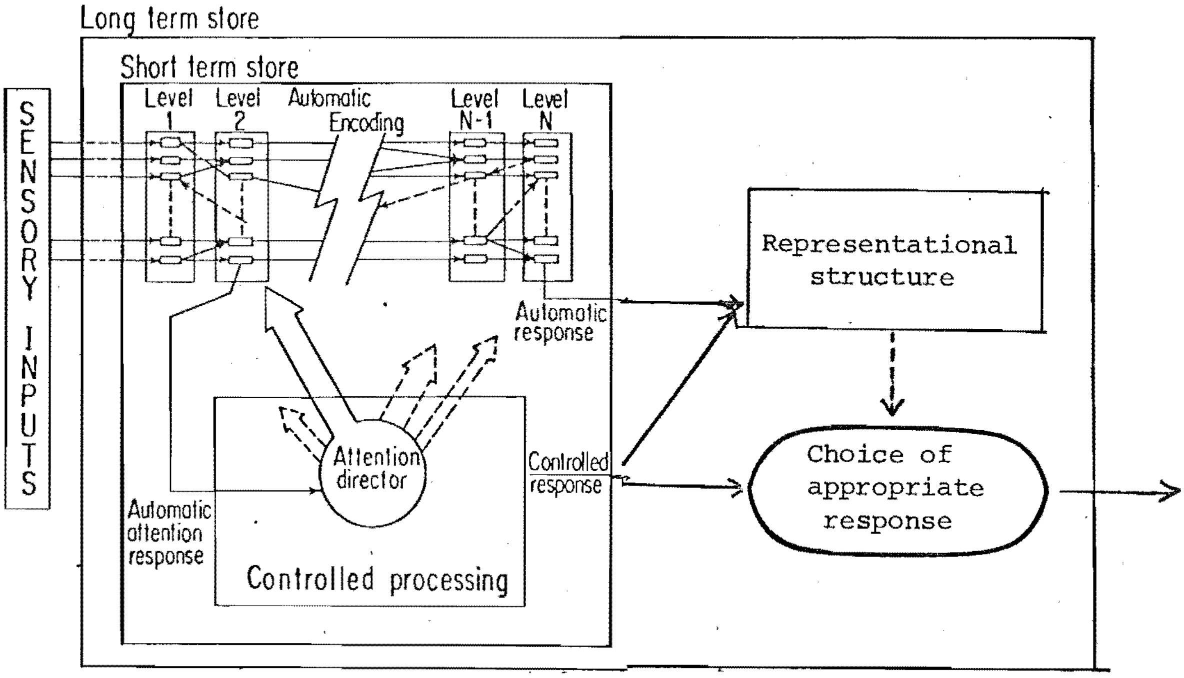


Figure 4.3 The model (theory[1]) shown in figure 4.2 modified to include the notion of representational structure. Response derivation is shown as the choice of the appropriate (experimenter-defined) response which will be based on the controlled (implicit) response.

Figure 4.3 shows a modified version of the automatic and controlled processing model. A representational structure of the stimuli will be built up during the experiment while the available responses (for example, the numbers one to seven) may (possibly) be viewed as logogens which act as response strength accumulators. The response (in this view of response derivation) will be made once the response threshold exceeds the threshold for a particular logogen.

1. In terms of the distinction between theories and models made in Chapter One, most of the 'models' cited in this chapter are actually theories. They are called models here so as to conform with the terminology used in the original sources.

Multidimensional Scaling (MDS) and Representational Structure

The purpose of multidimensional scaling has been outlined previously in the following manner:

"Multidimensional psychological scaling serves the purpose of mapping *percepta* or images, the subjective equivalents of *stimuli* (objects, persons, comcepts, etc.) as *(stimulus) -points* into a fictitious model-space, isomorphic to the real vector space R with appropriate metrics. This space is usually called *subjective* or *psychological space*, its dimensionality n equals the minimum number of attributes necessary for a complete but nonredundant description of the percepts studied."

—Micko and Fischer (1970, p.118).

The possibility that a multidimensional scaling solution may represent psychological space allows the researcher to use it as an estimate of psychological structure which would not otherwise be accessible to him/her. Within the framework of this thesis, a multidimensional scaling solution of similarities data is regarded as an estimate of the participants' representational structure. According to this model, the participant derives his/her similarity responses from the representational structure while the experimenter uses multidimensional scaling analysis (MDS) in an attempt to recover that same representational structure of the stimuli which is presumably implicit in the pattern of similarity responses.

An alternative model would be that responses are chosen independently of representational structure. The responses would then be made to satisfy the requirements of the experiment while the representational structure would be the way the participant organises the stimuli quite independently of the need for choosing the responses which the experiment requires. The first model will be assumed below as it allows the use of multidimensional scaling solutions to estimate representational structure. Despite this pragmatic approach, there may be situations where the responses are not derived from the representational structure.

Ecological Approaches to Perceptual Modelling

The framework of this thesis (cf. chapter one) is oriented around the issue of how man obtains and acts upon knowledge of his environment. This appears to be far more closely related to the 'ecological approach' of Gibson (1966) and others (Mace (1977) gives a brief summary of the position) than to the type of model of indirect perception so far being considered. Turvey (1977, 1978) has strongly criticised what he sees as the 'indirect realism' of most cognitive theorists:

"Presumably, the goal of visual processing theory is to isolate and characterize that which is most eminently and directly responsible for our perceptual knowledge. In the view of indirect realism, the candidates for this honor are patently the postulated links in the internal chain of epistemic mediators from retinal image to perceptual experience . . . But the view of direct realism promotes a very different roster of candidates. They are, most obviously, the complex, nested relationships in the dynamically structured medium surrounding the observer that are specific to the properties of the environment in which he or she acts."

—Turvey (1977, p.86).

Despite the persuasive arguments of Turvey and others, the ecological attitude to research and theory on visual processing has yet to produce much in the way of quantifiable theory. Consequently the present approach will continue to use the assumptions currently necessary for the quantifying of models, although the philosophical difficulties inherent in such an approach (as pointed out by Turvey, 1977, 1978) are recognised.

Task Demands and Selective Attention

One of the key features in the type of theory represented in Figure 4.3 is controlled processing. The experiments to be reported in the remainder of this chapter investigate the relationships between controlled processing, attentional strategy, and task demands. The effect of task demands on the use of controlled versus automatic processing has been shown previously (Schneider and Shiffrin, 1977).

The work of Shiffrin and Schneider (1977) suggests that task demands will have their effect either by forcing the participant to switch attentional strategies within the controlled processing mode, or else by forcing him/her to switch between the controlled and automatic processing modes. Since the present thesis is concerned with building a framework for perceptual learning, it will be looking at the preasymptotic response behaviour to an originally novel set of visual stimuli, and thus it can be assumed that participants will generally be using controlled processing.

Following the rationale of the preceding discussion, it is hypothesised that the changing of task demands in experiments using the Walsh stimuli will result in changes in attentional strategy. The remainder of this chapter will provide the necessary information on the effects of changes in attentional strategy so that it will then be possible to quantify the effects of different task demands in Chapter six.

Separability and integrality of Stimulus Dimensions

For the last twenty-five years or so, Garner and his colleagues have been concerned with stimulus structure and its effect on perception (Garner, 1974 , 1978). The bulk of this work on stimulus structure (or, at least, the part of that work which is relevant to the present thesis) can be summarised in the following seven axioms.[1] which are taken directly from Garner (1974 , pp. 120–121):

1. The structure of stimulus sets may be based upon similarity relations between stimuli, or it may be based upon dimensional relations between stimuli.
2. Stimulus dimensions that produce sets in which similarity is important are termed integral. Those that produce sets in which dimensional structure is important are termed separable.
3. In direct similarity scaling, integral dimensions produce interstimulus relations with a Euclidean metric; separable dimensions produce interstimulus relations with a city-block metric.
4. In perceptual classification, stimulus sets defined by integral dimensions are classified primarily in relation to similarities; sets defined by separable dimensions are classified in relation to dimensional structure.
5. In perceptual classification, dimensional preferences or saliences exist only for separable dimensions.
6. Manipulation of relative discriminabilities of dimensions has little effect on the dimensional preferences exhibited with separable dimensions, while almost completely determining classification with integral dimensions.
7. Both similarity and dimensional structure of sets of stimuli exist for both integral and separable dimensions. However, with integral dimensions, the primary structure is similarity in the sense of distance, while the dimensional structure is based upon a more derived, cognitive process. On the other hand, with separable dimensions, the primary structure is dimensional, and the similarity structure is based upon a more derived, cognitive process.

Torgerson (1965) originally suggested that similarity is not a unitary concept, but has a different meaning depending on whether the stimuli have analyzable (separable) or unitary (integral) dimensions. Point three above summarises a theory about the effect of separability-integrality on the existence and the metric of the appropriate multidimensional spaces fitted to pairwise similarities data. (This follows from the work of Attneave, 1950; Shepard, 1964; and Hyman and Well , 1967, 1968).

Since subsequent experiments will be using pairwise similarity judgements and multidimensional scaling, it may be important to know whether the Walsh stimuli have integral or separable dimensions.

1. Although these points appear to have been presented as conclusions by Garner, there is not sufficient evidence in support of them to suggest that they are more than axioms. Point three, for instance, is directly contradicted by the results given later in this chapter.

In terms of their construction, as multiples of underlying Walsh functions, the Walsh stimuli would appear to be highly analysable. Participants have no trouble in counting the number of horizontal or vertical stripes in order to determine the row and column sequencies. Further evidence for the analysability of the Walsh stimuli comes from the sorting tasks which are described in chapter eight of this thesis. According to Garner (1974) the sorting of analysable stimuli will be in terms of dimensions (in the case of the Walsh stimuli, it might be putting all the stimuli with the same number of stripes into one pile) whereas unitary stimuli will be sorted in terms of similarities. The preponderance of rule-based sorts obtained in the experiment described in chapter eight also suggests that the Walsh stimuli are analysable.

Assuming that the Walsh stimuli are in fact analysable (separable), axiom seven above indicates that the primary structure should be based upon a more derived, cognitive process. In such a case:

"... (T)he metric of the subjective space will be a two-dimensional Minkowski power metric, intermediate between the city-block metric and the Euclidean metric, if a pair of independent subjective attributes is mainly although not exclusively considered in the determination of overall subjective similarity."

—Micko and Fischer (1970, p.126).

The following two experiments sought to investigate the effect of selective attention instructions on the Walsh stimuli, both in terms of the features which were used to make the similarity judgements, and in terms of the metric of the underlying subjective space. Apparent changes in the metric of the subjective space have previously been considered by Hyman and Well (1967, 1968) who showed by an indirect argument that the Minkowski exponent of the colour space decreases if the participant's attention is drawn towards particular colour variables.[1]

Splitting the Stimuli into Subsets

The selection of 35 of the Walsh stimuli for use in further experimentation was described in chapter three. A complete pairwise comparisons experiment for these stimuli would involve (35.34/2) comparisons, that is, a total of 595 comparisons. The 35 stimuli were split into three subsets of 15 for the purpose of experimentation using pairwise similarity judgements. These three subsets were overlapped as shown in Figure 4.4 so

1. Wender (quoted in Fischer and Micko, 1972) made the Minkowski exponent (an exponent of one indicates city-block space while two indicates Euclidean space) increase by shortening the exposure time of highly analysable geometric figures. However, this result is mathematically ambiguous because of the possible conjugacy of MDS solutions (where one solution is for a metric with an exponent of between one and two while the other solution has an exponent greater than two), and the general robustness of Euclidean solutions (see Shepard, 1974, pp. 406-408 for a discussion of both these points).

that an estimate of the total structure for the 35 stimuli could be made.

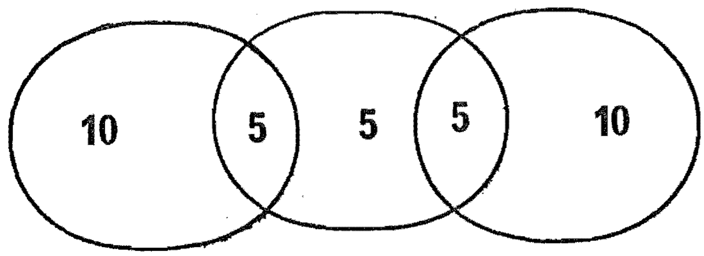


Figure 4.4 A schematic representation of the way in which the three subsets of stimuli overlap.

Set one consisted of 15 of the 35 stimuli which did not have underlying row or column sequencies of four or eight, and which were not on the negative diagonal (bottom left-top right) of the sequency-ordered matrix of Walsh stimuli (Appendix A, Figure A.1).

Set two consisted of stimuli which were either located on the negative diagonal of the sequency-ordered matrix or which had at least one underlying sequency of four or eight.

Set three consisted of five of the stimuli from set one, five from set two, and the five stimuli remaining which had not been placed in either set one or set two. Table 4.1 identifies the Walsh stimuli which belonged to the three sets.

Experiment SIMS1

Method

12 participants each attended a forty minute session where they were presented with all possible pairs (105) of the 15 set one stimuli in random sequence. They were asked to rate each stimulus pair in terms of similarity on a seven point scale (1 = dissimilar, 7 = similar).

Results

The first analysis used POLYCON (Young, 1973) to scale multidimensionally the 15 stimuli of set one using the city-block model. All 12 participants were used as replications in the analysis with a two-dimensional solution being fitted. (The analysability of the Walsh stimuli into row and column sequency suggested that a city-block model should give a good fit to the data.) The model did not provide a good fit with Kruskal's (1964) stress formula two,[1] having a value of .614. To check whether this poor fit was due to individual differences, the analysis was re-run with the data from a single participant (This participant

1. All the stresses reported here are calculated using Kruskal's stress formula 2.

Set One	Set Three	Set Two
3		
5		
6		
11		
13		
14		
15		
22		
38		
39		
2	2	
7	7	
21	21	
23	23	
47	47	
	46	
	48	
	55	
	56	
	30	
	4	4
	10	10
	20	20
	24	24
	64	64
		8
		12
		16
		19
		28
		29
		31
		32
		37
		40

Table 4.1 Stimulus composition (given as Walsh stimulus #s) of the three stimulus subsets.

was chosen on the basis of the INDSCAL analysis reported below, that is, the participant at the centre of the subject space).

The solution is shown in Figure 4.5 (it had a stress of .432).

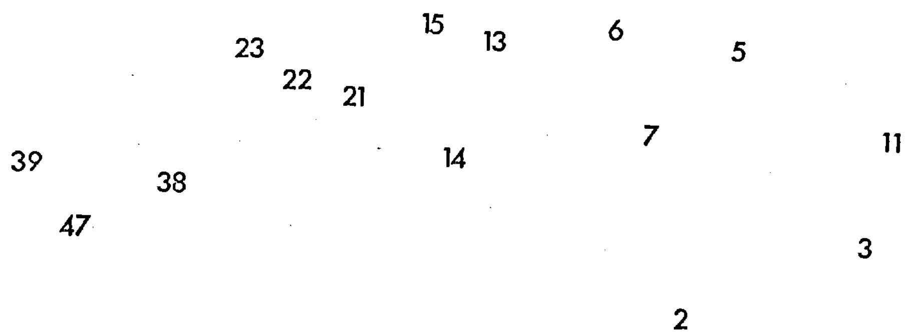


Figure 4.5 City block MDS solution (in two dimensions) for one of the participants from SIMS1.

The same analysis was then run except that this time the Euclidean model was fitted. The solution is shown in Figure 4.6 (it had a stress of .437) and is similar to the city-block solution except that the second dimension is more spread out. The solutions shown in figures 4.5 and 4.6 had similar stress. The relatively good fit of the Euclidean metric in comparison with the city-block metric does not necessarily imply that a Euclidean metric is reasonable:

"... (P)urely Euclidean solutions can be surprisingly robust in the face of certain kinds of rather marked departures from the assumed Euclidean metric."

— Shepard (1974, p.407)

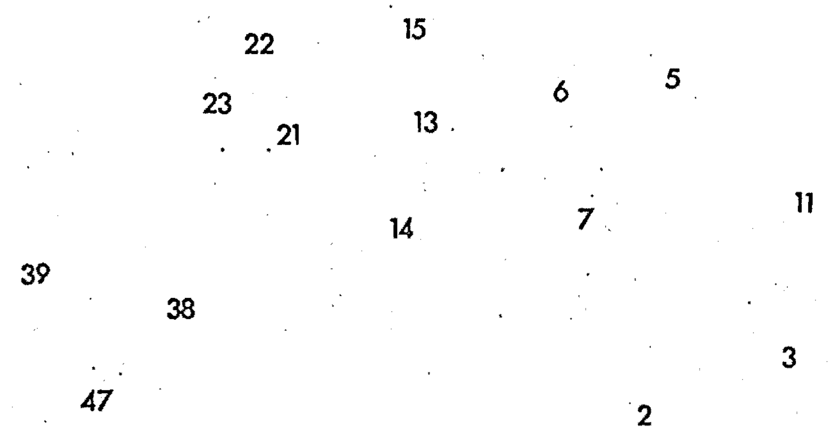


Figure 4.6 Euclidean (2–D) MDS solution for the same participant as in Figure 4.5.

However, Wiener-Ehrlich (1978) regressed the dissimilarity judgements of one-dimensional judgements on the mean dissimilarity ratings on each of 120 trials (she used 16 stimuli) with a set of analysable stimuli. She looked at the percent variance accounted for by a regression equation corresponding to the Euclidean metric as against an equation corresponding to the city-block metric. The results of this analysis indicated that the city-block metric provides only a marginally better approximation to subjects' rule of combination than the Euclidean metric. A similar analysis was carried out on the SIMS1 data and will be reported in a later section of this chapter.

Summary of First Analysis

Multidimensional scaling of the SIMS1 results using both the city-block and Euclidean models produced similar results. This lack of evidence for a better fit for the city-block metric when applied to analysable stimuli contradicts the third of Garner's (1974) seven points (which were outlined above).

The similar configurations in Figures 4.5 and 4.6, along with the relatively low stress value of the Euclidean solution suggest that the Euclidean model can be used as a good first approximation to the best fitting distance model within the Minkowski family of metrics. Consequently, INDSCAL (Carroll and Chang, 1970) was used to account for possible individual differences in the data.

INDSCAL Analysis

INDSCAL solutions in two and three dimensions were fitted to the SIMS1 data. Since the correlations between computed scores and the original data were similar for both dimensionalities, the two-dimensional solution was selected as providing the most parsimonious representation of the results. Figure 4.7 shows the INDSCAL solution space for SIMS1.

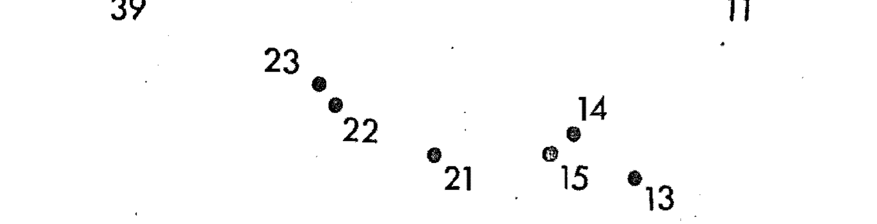


Figure 4.7 Two-dimensional INDSCAL solution space for the SIMS1 results.

It can be seen that this solution is generally similar to the POLYCON-derived solutions shown in Figures 4.5 and 4.6. Dimension one of the SIMS1 INDSCAL solution appears to be the reverse of either two-dimensional sequency (the correlation between 2-D sequency and dimension one is $-.970$) or the complexity score ($r = .966$).
Dimension two is not so easy to identify with only one feature having a significant correlation with it ($p < .05$ for a 1-tailed test, using the Pearson Product-Moment table of critical values in Fisher and Yates, 1963). This feature was component one (Feature 24) and had a correlation of $-.628$.

Experiment SIMWR

The next experiment sought to quantify the effect of selective attention instructions on judgements of similarity between Walsh stimulus pairs.

Method

12 participants (none of whom had participated in SIMS1) attended an experiment which was identical to SIMS1 apart from the task instructions. They were asked to judge the SIMilarity With Respect (SIMWR) to complexity of each slide pair and to rate this similarity on a seven-point scale[1] (identical to the rating scale used in SIMS1). The complexity of a given stimulus was defined as the time the participant thought would be necessary to memorise that stimulus (this is analogous to the definition of complexity used in experiment E2 of Chapter two).

Results

INDSCAL solutions in two and three dimensions were fitted to the SIMWR data and the two dimensional solution was selected for further study (cf. the INDSCAL analysis for SIMS1). Figure 4.8 shows the INDSCAL solution space for SIMWR.

1. The similarity scales used in Chapters four, five, and six were all loosely anchored in the following manner.

The participant was told to rate seven if the pair were completely similar, one if they were completely different, and four if (s)he was uncertain whether the pair were more similar than different, or not.

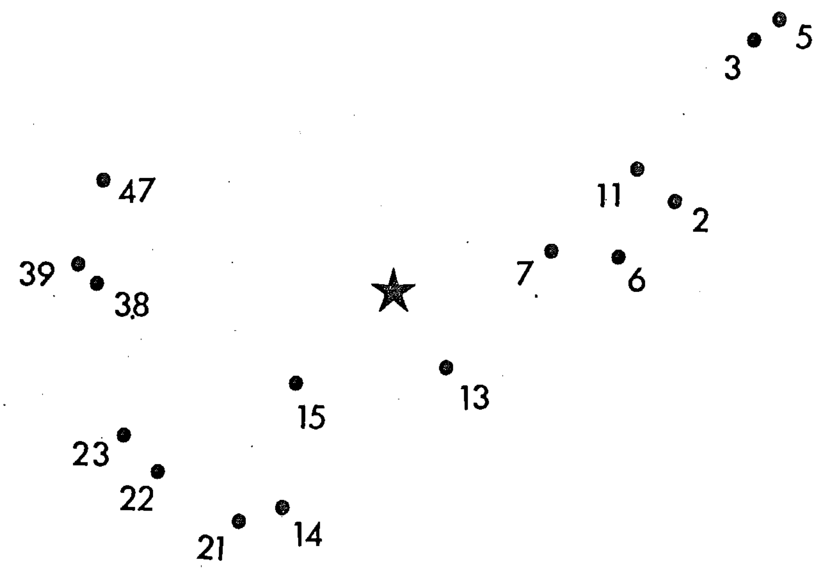


Figure 4.8 Two-dimensional INDSCAL solution for the SIMWR results.

It can be seen that the INDSCAL solutions for SIMS2 and SIMWR are roughly the same. Figure 4.9 shows the displacement vectors for each of the 15 stimuli; these represent the change in the SIMS1 configuration due to the selective attention instruction. Figure 4.9 does not show any consistent pattern which can be ascribed to a selective attention effect.

Identifying the INDSCAL Dimensions

The correlation of dimension one with two-dimensional sequency (r = .656) in the SIMWR configuration was much smaller than for the corresponding SIMS1 configuration, while the correlation between dimension one and the empirical measure of complexity (defined above and in chapter two) was –.792, and the correlation of dimension 2 with the empirical measure of complexity is –.929.

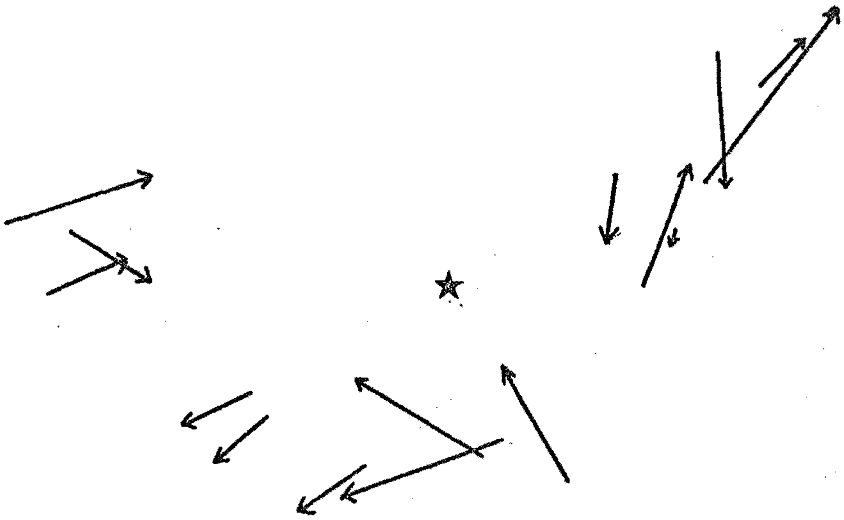


Figure 4.9 Vectors represent the changes in position of each of the 15 stimulus point between the SIMS1 (tails) and SIMWR (heads) solutions, with the direction of supposed attentional effect represented by the arrows.

Implications for Similarity Modelling

Two dimensions appeared to underlie the SIMS1 results (these can be approximated by two-dimensional sequency and component one) whereas only one dimension (perceived complexity) appeared to underlie the SIMWR results.

If the INDSCAL analyses have extracted the features underlying the similarity judgements in SIMS1 and SIMWR, similarity models based on two-dimensional sequency, component one, and complexity should be able to account for the similarity ratings. The INDSCAL method incorporates the following distance model (Carroll and Wish, 1974, p.59):

$$d_{jk}^{(i)} = \left(\sum_t w_{it} \, (x_{jt} - x_{kt})^2 \right)^{1/2} \qquad (4.1)$$

Where $d_{jk}^{(i)}$ is the distance between stimuli j and k for the i^{th} participant, x_{jt} is the co-ordinate of stimulus j on the t -th dimension and w_{it} is the salience of dimension t for participant i. An analogous regression model is given below:

$$(d_{jk}^{(i)})^2 = \sum_t w_{it} \, (x_{jt} - x_{kt})^2 \qquad (4.2)$$

where x_{jt} is the value of feature t for stimulus j and the regression equation is separately fitted for each participant i. Thus the Euclidean model is equivalent to a regression where the predictors are the squares of the feature differences between the stimulus pair in each trial, and the criterion will be the square of the distance between the stimulus pair in each trial. Using the same notation as above the city block model is as follows:

$$d_{jk}^{(i)} = \sum_t W_{it} \cdot \text{abs}\,(x_{jt} - x_{kt}) \qquad (4.3)$$

Thus the difference between the INDSCAL and multiple regression approaches involves the use of scaling to derive the underlying dimensions as against the selection of subsets of a priori features as predictors. Interpretation of the INDSCAL results suggested the use of three features in the regression analysis (i.e., two-dimensional (2–D) sequency, component one, and complexity). There are a total of $2^3 - 1$ (7) non-empty subsets of these three features which may be used as regression model equations. The regression equations derived from each of the seven subsets of features were fitted to the results of each of the SIMS1 participants using the city-block form of regression (4.3). The similarities were multiplied by negative one to convert them to distances for the regression analysis. The model chosen as representing a given participant was that regression equation which maximised the squared multiple correlation between the predictor and the criterion while using the smallest possible number of predictors. The decision as to whether an extra predictor should be included in the equation was based on a test of the hypothesis that the two regression equations (one with and one without the extra predictor) produced equivalent multiple correlations (had the same predictive effect). The test statistic [1] used is described in Afifi and Azen (1972, p.132) and is shown below:

$$F = \frac{n - l - 1}{l - h} \cdot \frac{R^2_{y.l} - R^2_{y.h}}{1 - R^2_{y.h}}$$

where F has an F distribution with $l - h$ and $n - l - 1$ degrees of freedom. l represents the number of predictors in the larger equation while h is the number of predictors in the smaller (in the sense of fewer predictors) equation. n = the number of trials (105). $R_{y.l}$ is the multiple correlation between the criterion and the set of l predictors; $R_{y.h}$ is the multiple correlation between the criterion and the set of h predictors.

Table 4.1 gives the fitted models for 10 of the SIMS1 participants,[2] including the squared multiple correlation and the squared INDSCAL correlation.

1. The F statistic is used to test H_0: $R_{y.h} = R_{y.l}$

2. Two participants were omitted because of practical problems associated with the retrieval of their data. The results of one of the participants will not be used in the corresponding SIMWR analysis for the same reason.

Participant Number	Sequency W_1	Component One W_2	Complexity W_3	R^2	INDSCAL R^2
1	.57	.47		.710	.650
2		.38	.60	.701	.702
3	1.00			.614	.765
4			1.00	.394	.517
5		.33	.55	.563	.654
6		.48	.42	.578	.570
7	.27	.14	.47	.582	.622
8		.30	.51	.492	.613
9			1.00	.212	.457
10		.62	.22	.548	.624

Table 4.1 The best fit regression model beta weights and variance accounted for, for 10 of the SIMS1 participants. (The INDSCAL variance accounted for is also shown).

A comparison of the squared multiple correlation and the squared INDSCAL correlation appears to show that the INDSCAL solution generally gives a better fit (average percent variance accounted for per participant was 62% for INDSCAL and 54% for the regression model). The correlation between the variances accounted for by the INDSCAL and regression models was .83 ($p < .005$) which reflects the similarity between the models (INDSCAL may be regarded as a combined scaling – multiple regression procedure.)

The procedure described above was repeated using the SIMWR results. Table 4.2 gives the fitted regression models for 11 of the SIMWR participants.

Participant Number	Sequency W_1	Component One W_2	Complexity W_3	R^2	INDSCAL R^2
1	.65	.21		.560	.597
2	1.00			.483	.642
3		.36	.37	.390	.610
4	1.00			.601	.756
5	1.00			.706	.701
6			1.00	.461	.697
7			1.00	.505	.771
8	.42	.25	.20	.519	.700
9			1.00	.548	.708
10		.22	.65	.597	.679
11			1.00	.575	.719

Table 4.2 The best fit regression model beta weights and variance accounted for, for 11 of the SIMWR participants.

The average fit of the regression models was the same as for SIMS1 (54% of the variance) whereas the INDSCAL solution accounted for an average of 69% of the variance in SIMWR. Table 4.2 indicates that single-predictor regression models provided the best account of seven of the eleven participants' results, whereas the corresponding proportion for SIMS1 was only three out of ten. It thus appears that the selective attention instruction in SIMWR had the effect of making most participants judge similarity with respect to either 2–D sequency[1] or complexity (feature 27). This apparent reduction in the number of parameters needed in fitting the similarity model (which is represented by the family of seven regression equations fitted to the SIMS1 and SIMWR results) is a limiting case of the parameter imbalance which was found by Gregson (1972) as a response to directed selective attention.

Some Additional Points

Kruskal and Wish (1978, Appendix B) have described the "horseshoe" phenomenon which sometimes occurs in a two-dimensional MDS configuration. It consists of a nearly one-dimensional configuration which has been bent around into a horseshoe shape. The horseshoe phenomenon is apparent in the 2–D INDSCAL solutions for SIMS1 and SIMWR (figures 4.7 and 4.8). The high correlation of both the SIMWR solution dimensions with complexity suggests that complexity is the only dimension underlying the SIMWR results. The bending of this underlying dimension into a horseshoe may possibly be explained as the utilisation of the second dimension to account for residual variance.

Wiener –Ehrlich (1978, experiment one) used regression models of the same form as equations 4.2 and 4.3 to assess which metric (out of the city-block and Euclidean metrics) gave the better fit to results obtained using analysable/separable stimuli. She found (using pooled results) that the two forms of regression equation gave approximately the same fit to the results. The SIMS1 results were reanalysed to check this finding with individual (rather than pooled) results. Regression equations (derived from the subset of features which had previously given the best fit using the city-block model) were fitted to the results of each of the SIMS1 participants using the Euclidean form of regression (4.2). The variance accounted for was lower than in the equivalent city-block regression analysis for each and every participant. An average of only 40% of the variance was accounted for per participant as against the 54% average for the previous analysis. Thus the SIMS1 results were fitted appreciably better by the city-block metric whereas Wiener-Ehrlich (1978, experiment one – using squares varying in size and brightness) found that the Euclidean metric gave a better fit. The

1. 2–D sequency is similar to measures of complexity which have been used previously (e.g., the uncertainty measure used by Dorfman and McKenna, 1966).

discrepancy between the results of Wiener-Ehrlich (1978) and those obtained here will not be discussed further, however, as no attempt was made to check the validity of the additive difference model (which is implied when using multidimensional scaling with Minkowski metric) for the similarity judgements of the Walsh stimuli.

Summary

The present chapter sought to develop methods for quantifying the effects of task demands on psychological judgement. The first section expanded the perceptual-cognitive theory outlined in Chapter one to include the effects of task demands on attentional strategy. The second section considered two experiments; one with, and one without selective attention instructions. (Apart from this one difference, the experiments were identical). It was shown that the implied metric of the best fitting MDS solution was not a sensitive measure of selective attention. Separate regression models were fitted to each of the participants in the two experiments. It was found that models with three or fewer predictors were able to account for between 40% and 70% of the variance in similarity responding for all but three of the (retained) participants. INDSCAL analysis was also used, and found to account for a large proportion of the experimental variation. (An average of 68% of the variance for each participant in SIMS1.)

The results presented in this chapter raise a number of substantive issues which are outside the scope of the present thesis.

Firstly, the apparent robustness of the Euclidean metric to analysable stimuli found both here and in previous studies suggests that a Monte Carlo simulation study of the effect of data structure (linear verses quadratic combination rules, for instance) on the apparent metric of multidimensional representations would give researchers guidance (which is presently lacking) in interpreting the type of results being considered here. The work of Spence (1970) provides a good example of the type of study required.

Secondly, the relationship between MDS and regression requires some clarification. In the present study if was found (for the SIMS1 results) that the Euclidean metric was adequately able to account for the results when using MDS (INDSCAL) but was inferior to the city-block metric when using regression analysis.

Thirdly, the presence of individual differences in similarity judgements and the important role that attentional strategy has in current cognitive modelling suggest that balanced repeated measures designs are necessary to characterise the sequential interaction between task demand effects on selective attention and individual differences using a range of stimulus materials. This has been anticipated by Gregson (1969) with respect to changes in task demand due to changes in the stimulus set but has yet to be studied using other methods of changing task demands.

Fourthly, the evidence presented here for individual differences in the actual set of predictors in the best fit models of similarity judgement suggests that current approaches to individual differences scaling may be failing to allow for an important component of individual variation. As Lingoes and Borg (1978) have noted, neither the IDIOSCAL family of procedures (Carroll and Wish, 1974) nor recent non-metric applications such as ALSCAL (Takane et al., 1977) add anything to the dimensional salience[1] model which might provide information regarding individual differences of a more complicated nature. The rotations and translations available in PINDIS (Lingoes and Borg, 1978) do not allow for the simple possibility that different people may perceive different dimensions either. In view of this it is suggested that more precise modelling of the response behaviour of individual participants is necessary to account properly for individual differences in the perception of stimuli which do not have a unique, explicit, and culturally defined dimensional structure.

1. Dimensional salience here is the weighting of MDS derived dimensions to account for individual differences. This should be distinguished from the process of separately fitting a model (or models) for each individual which is the approach advocated here.

CHAPTER V

It was suggested in chapter one that similarity judgements may be an important part of perceptual learning. This notion of similarity as a cognitive process should be distinguished from the use of similarity judgements as a means of estimating the psychological structure of a stimulus set. This chapter will derive empirical estimates of the psychological structure of the Walsh stimuli, as well as clarifying the distinction between similarity judgement as a data collection technique and similarity judgement as part of cognitive processing in general. In terms of the framework being used in this thesis, similarity judgements can be used to estimate representational structure. Some researchers have regarded the elaboration of representational structure as being the main reason for collecting similarity data:

"I shall take the primary purpose of the analysis of such a triangular matrix of similarity measures to be the achievement of a concise, invariant, and assimilable representation of the essential pattern of structure that lies more or less hidden in the given array of numerical data. By the achievement of the 'invariant' representation of the 'essential' structure, here I mean to exclude representations that are heavily influenced by arbitrary features of the data: representations, for example, that change appreciably when the data are subjected to seemingly permissible transformations."

— Shepard (1974, p.374).

This approach ignores a large amount of psychological research (see Gregson, 1975, for a review) which shows that similarity judgements are generally quantifiable as decision processes which can be approximated by a variety of similarity models.

Shepard's approach appears to assume the existence of a single MDS solution space which can be regarded as an appropriate representation of the psychological space. The results of experiments to be reported below, as well as studies such as that of Homa, Rhoads and Chamblis (1979) show that there are changes in the MDS solutions, derived from different experiments, run at different times, which appear to be due to learning effects and variations in the experimental task. This appears to preclude the possibility of identifying a single representational structure (for a given stimulus set,) which is independent of task and learning effects, unless an overlearned stimulus set with a completely unambiguous perceptual and conceptual structure is used.

Perceptual and Cognitive Similarity

The implications of models of similarity for perceptual-cognitive theory have been recognised in a number of recent papers (Tversky, 1977; Ortony, 1979).
The role of similarity as an important part of the incremental process of knowledge acquisition is also receiving attention:

"... (I)t should be noted that similarity plays a dual role in theories of knowledge and behavior: it is employed as an independent variable to explain inductive processes such as concept formation, classification, and generalisation; but it is also used as a dependent variable to be explained in terms of other factors ... Similarities are constantly updated by experience to reflect our every-changing picture of the world."

— Tversky and Gati (1978, p.98).

However, despite the recognition that similarity is a part of perceptual-cognitive processing as well as a phenomenon to be studied in its own right, there have been few attempts to precisely identify the role of similarity in perceptual learning. The discussion below and the experiments following it represent an attempt to account for similarity judgements within the framework of this thesis (outlined in Chapters one and four). This is a first step towards replacing the notion of similarity as either an independent or a dependent variable with the theoretical notion that similarity is part of a perceptual-cognitive process.

Similarity judgement tasks can be distinguished in terms of the amount and type of processing they require. Previous studies using similarity judgements have tended to collect similarities data under standard conditions for the purpose of quantifying *the* psychological structure of the stimuli as perceived by the participant (Shepherd, 1974). In general, the stimulus pairs are presented simultaneously with no significant time constraint.[1] In the present view such tasks measure perceptual aspects of stimuli in that the stimuli may be directly compared one with the other. This is not to say that cognitive factors such as familiarity are not involved, but these factors would have no more effect than they would have in any act of perception. Cognitive similarity refers to similarity judgements made between stored representations of stimuli or between concepts which are not perceivable, even in principle. An example of this latter type of judgement is the rating of pairs of countries in terms of their similarity.

Asymmetry in Similarity Judgements

The fitting of models involving distance metrics (these include most forms of MDS) utilises the assumption that the distance from point x to point y, d(x,y), is the same as the distance from point y to point x, d(y,x). This assumption is not always valid for psychological data. Krumhansl (1978, pp. 451-452) reviews the empirical evidence for asymmetry. Tversky (1977) has characterised similarity as a relationship which has a subject and a referent. According to Tversky and Gati (1978) the choice of which of two geometrical stimuli will be the subject and which the referent depends on the relative

1. Many studies use the presentation of stimulus pairs in booklet form where the participant may work at his/her own speed.

saliences of the two stimuli which in turn depend on 'goodness' of form and 'complexity'.

Tversky and Gati (1978, study three) asked their participants to rate (on a 20-point scale) the degree to which the figure on the left was similar to the figure on the right, for each of 120 pairs of geometrical figures. They found differences when the left and right figures in each pair were reversed, but no account of the size or generality of these differences across participants was given. (The differences were assessed between two groups of 66 participants using t-tests). While asymmetry in similarity judgements has important implications for the type of similarity model which will be appropriate (Tversky, 1977), the previous research designed to demonstrate its presence has not yet shown that asymmetry has a significant effect when the stimuli in a pair are presented simultaneously without predisposing instructions (or a set brought to the task by the observer) as to which is the subject and which the referent.

In the present view asymmetry is more usefully seen as an experimental task demand effect. Unless there are other task demands (such as delayed presentation of the second stimulus in the pair) the participant can be expected to use the most familiar (memorisable, prototypical, salient) stimulus in the pair as a frame of reference in making the similarity judgement. If there is a delay in the presentation of the second stimulus, then the first stimulus will tend to become the frame of reference. According to the above argument the best way to investigate asymmetry effects may well be to introduce a tradeoff between time delay of the second stimulus and its degree of familiarity (in comparison with the familiarity of the first stimulus). It would be expected that the relative familiarity of the two stimuli should determine which one becomes the frame of reference at shorter time delays. As the Inter-Stimuli Interval (ISI) increases, the likelihood of the frame of reference being taken as the first stimulus should increase, with the effect of the relative familiarity of the stimuli becoming less and less important. This frame of reference effect will thus be an important consideration in the delayed similarities paradigm which is outlined in the next chapter.

The next section will explicitly relate similarity judgements to cognitive functioning in general.

Similarity within a cognitive framework

Sjöberg (1972) attempted to characterise the nature of cognitive, as opposed to perceptual, similarity. The conclusions of his study will not be considered here because they were not based on a proper specification or quantification of the two models proposed. However, Sjöberg's work is of interest because his implicitly set-theoretic notions (although not apparently recognised as such by Sjöberg) preceded the more comprehensive set-theoretic formulations of Gregson (1975) and Tversky (1977). Set-theoretic formulations of similarity are compatible with featural models of semantic memory (Smith, Shoben and Rips, 1974) and pattern recognition (Reed, 1973). Within the framework of this thesis, similarity judgements will be viewed as occurring with respect to the activated memory schemata and will involve a feature comparison process. (This position will be modified in the later discussion on delayed similarity judgements in Chapter six). When the stimuli are simultaneously and continuously available the features on which the comparison process is based will tend to be perceptual features. Where there are task demands which necessitate encoding and storage of one or both of the stimuli before the comparison, the features used will tend to be more abstract and cognitively-based. This distinction between perceptual and cognitive features is closely related to the physical, name-code distinction of Posner and Mitchell (1967).

In attempting to place similarity judgements within a cognitive framework it is necessary to give proper consideration to the parameters previously found to be necessary in cognitive models. Gregson's (1975) characterisation of core structure is compatible with the cognitive model outlined in chapters one and four of this thesis. The model considered here is equivalent to the most general core structure model (Gregson, 1975)

$$S = C_s^* (X, Y, i, M, t, \Psi i, Ti).$$

in words, this says that similarity S is given by the core structure C_s^* prediction for the stimuli X, Y; conditional on the individual i, the stimulus set M, the stimulus dimensions (components) t, the psychophysical function Ψi, and the response derivation Ti which is taken to be an order-preserving transformation. Similarity between a stimulus pair is assumed to be conditional on the participant, the stimulus context, the stimulus dimensions, the activated schemata, and the participant's response derivation process. Of the parameters of the similarity task mentioned above, only the nature of stimulus context has yet to be considered here, and this will be remedied in the next section.

The use of similarities within a cognitive framework has a bearing on the type of similarity model which will be appropriate:

". . . there is an important difference of emphasis between the studies reviewed in this chapter and, say, the multidimensional metric and non-metric spatial models of similarity to be considered later; here, similarity is taken as a judgment process in its own right, to be explained and quantified, rather than an intermediate step in data collection for the generation of maps of perceptual and judgment spaces. In the vector model experiments and related investigations, the problem of defining the 'internal psychophysics' or psychometrics of similarity, and related but different judgment processes involving stimulus comparisons — usually but not necessarily pairwise — is squarely faced rather than left obscure or defined out of existence."

— Gregson (1975, p.65)

The emphasis in this thesis is on the derivation of models of similarity from theoretical considerations rather than the identification of the model which fits best from a range of possible models (cf. Eisler and Ekman, 1959). This approach can be adopted for three reasons:

1. The aim of this thesis is to develop experimental paradigms which are appropriate within the general theory of learning presented in chapter one. Any model of similarity used should be compatible with this framework.
2. Previous research has shown that most similarities data can be fitted to approximately the same extent by a number of similarity models, some of which may differ markedly in their psychological implications (Gregson, 1975, 1976, 1979).
3. Theoretical distinctions between the models have become blurred since Krumhansl (1978) has shown that appropriately modified geometric models of similarity will be able to make the same predictions as those made by set-theoretic models.

In response to this ambiguity in the interpretation of similarities data I shall (pragmatically) adopt contrasting orientations to the analysis of, as against theorising about, similarity results.

(a) Similarity results will be assumed to be produced by some feature comparison process which is compatible with set-theoretic models of similarity.

(b) Analysis of similarity results will be done using the more tractable distance based methods of multidimensional scaling and regression analysis.

The remaining problem will then be to relate the implications of the distance based results in terms of the feature comparison process which is assumed to have generated the data. The work of Krumhansl (1978) indicates that such an ambivalent approach should be tenable.

Stimulus Context

The notion of stimulus context has been described by Garner (1974, lecture one). He considered what happens when you are shown the letter E:

"You do not know the total set of stimuli, much less the particular subset. Nevertheless, you immediately perceive that E is a member of a set. Perhaps you assume that (four) straight lines are the components of the stimulus . . . If you also assume that each line can exist or not exist, then there are (15) positive stimuli (excluding the one with all lines missing) and you have an inferred set of stimuli, the inference being based on the single stimulus.

Alternatively, you might have assumed that the set of possible stimuli consisted of all the letters of the alphabet, or just the subset of vowels. To illustrate how easily the single stimulus can lead to a different inferred subset, consider the same stimulus reversed so that it looks like the numeral 3. Now no inference of a set of letters would be made, but possibly an inference of a set of numerals would be made."

— Garner (1974, p.11).

The above quotation highlights the fact that it is the perceived set of possible stimuli (of which a given stimulus or stimulus pair is a subset) which provides the stimulus context rather than the actual set used by the experimenter. The particular properties of a stimulus pair may be expected to provide a unique stimulus context appropriate to that particular subset of two stimuli.

Thus the perceived properties of a stimulus will change as that stimulus is variously paired with other stimuli (Garner, 1974, p.10). Garner concludes that subjects selectively attend to the variable which best differentiates a stimulus pair. Such a conclusion is not justified in view of the small amount of evidence produced in support of it (Mavrides, 1970) but it serves as an hypothesis which deserves further investigation.

In summary, stimulus context in a given experimental trial has two components:

(1) the inferred set of stimuli, and

(2) the properties of the unique subset of stimuli presented in that trial.

Configurational Representation

Kruskal's (1977) view is that multidimensional scaling and clustering may complement each other inasmuch as scaling represents the information contained in dissimilar pairs while clustering represents the information contained in similar pairs. This view is supported by the work of Graef and Spence (referred to by Kruskal and Wish, 1978, p.46) who found that discarding only the smallest third or the middle third of the *dissimilarities* does not disturb the reconstruction of the multidimensional space, while discarding the smallest third of the *similarities* causes a severe degradation. It thus appears that the combined use of multidimensional scaling and clustering may be necessary to represent all the information in proximities data. An example of such an analysis is given by Morgan (1973) who used multidimensional scaling, hierarchical cluster analysis and non-hierarchical cluster analysis on a confusion matrix.

1. This is formally equivalent to stating that comparisons are in a supremum metric space, which as a generalisation has had no empirical support (Gregson, personal communication), even though as a process it can sometimes occur. The difficult question about propositions of this sort is "how do we specify the conditions under which they are sometimes true?"

The remainder of this chapter will describe a set of experiments designed to identify the perceived structure of the Walsh stimuli. Clustering analysis will be used to complement the information obtained from INDSCAL and an attempt will be made to relate the resulting configurational representation of the Walsh stimuli to the set of Walsh features derived previously.

Experimental Design

Three experiments were carried out in order to obtain a configurational representation of the 35 Walsh stimuli. One complete (105 trials) pairwise similarity experiment was carried out on each of the three subsets of stimuli. (The composition of these subsets was outlined in Chapter four). The experiments will be referred to as SIMS1, SIMS2, and SIMS3 respectively, with SIMS1 having been referred to already in Chapter four.

Participants

All three of the experiments (SIMS1, SIMS2, SIMS3) used second and third year psychology students. SIMS1 had 12 participants while SIMS2 and SIMS3 each had 11 participants.

Method

The participants attended a forty minute session where they were presented with all possible pairs (105) of the fifteen stimuli in the appropriate subset. (SIMS1 used Set 1, SIMS2 used Set 2, and SIMS3 used Set 3). The participants were required to rate the similarity for each pair on a seven-point scale. The end labels for this scale were "completely similar" and "completely different" as mentioned in Chapter four.

Results

The data obtained from each of the three experiments were analysed using INDSCAL. Two-and three-dimensional solutions were obtained. The three-dimensional solutions each accounted for about six percent more of the variance in responding than the corresponding two-dimensional solutions, although the latter provide a convenient way of displaying most of the information obtained in the analysis.

Figures 5.1 to 5.3 show the two dimensional INDSCAL group space solutions derived from each of the three experiments while Tables 5.1 to 5.3 give the dimensional saliences for the participants. An examination of Figure 5.1 (and Figure 5.2) shows the possible existence of the "Horseshoe" phenomenon (Kruskal and Wish, 1978, Appendix B). Kruskal and Wish suggest that when the horseshoe phenomenon occurs, position along the horseshoe can be interpreted as another underlying dimension along with the directions in the space. Referring back to chapter four, the two-dimensional INDSCAL solution for SIMWR shows a

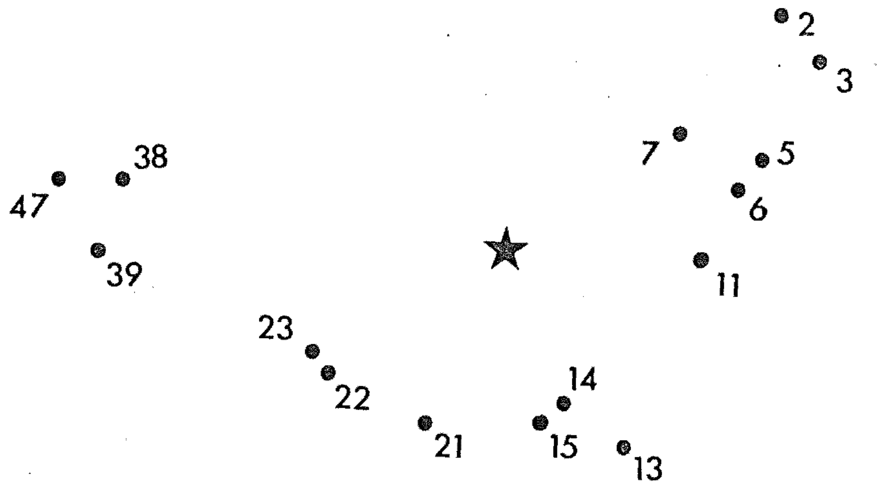


Figure 5.1 Two-dimensional INDSCAL solution for the SIMS1 results.

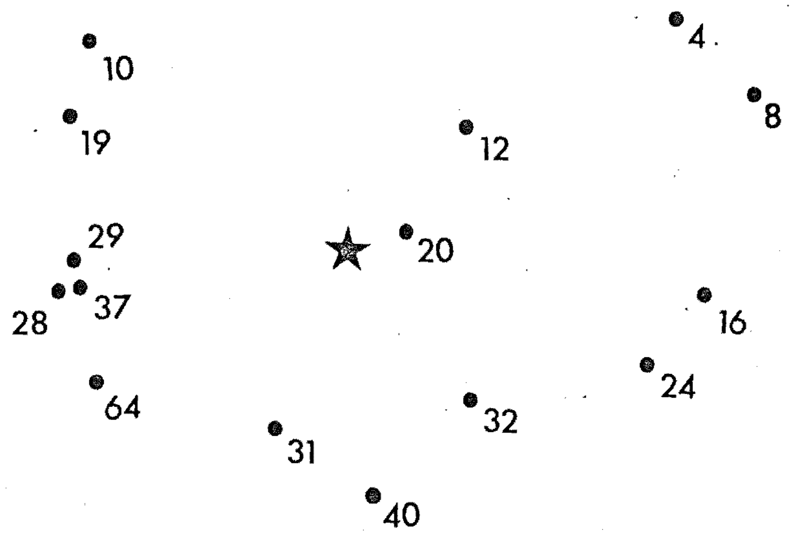


Figure 5.2 Two-dimensional INDSCAL solution for the SIMS2 results.

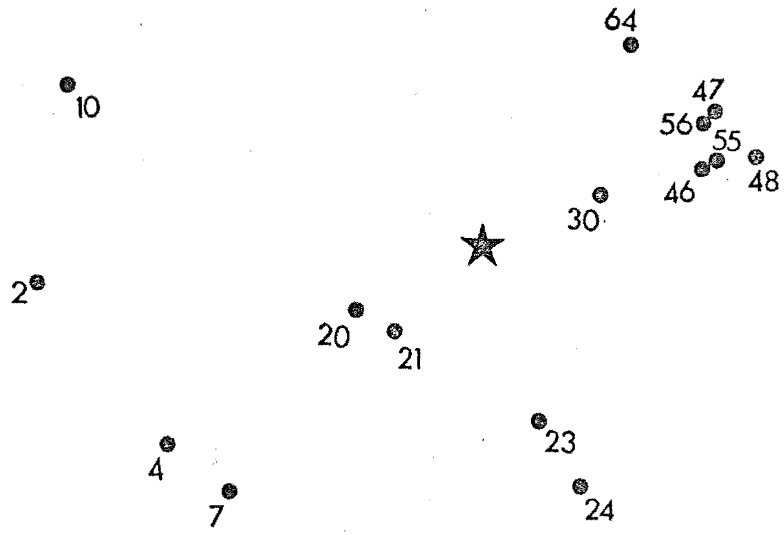


Figure 5.3 Two-dimensional INDSCAL solution for the SIMS3 results.

TABLE 5.1: Subject space dimensional weights for the SIMS1 Solution.

	Dimension 1.	Dimension 2.
1	.779	.249
2	.733	.354
3	.444	.474
4	.643	.396
5	.599	.467
6	.590	.364
7	.720	.313
8	.700	.356
9	.663	.375
10	.761	.208
11	.856	.120
12	.639	.350

TABLE 5.2: Subject space dimensional weights for the SIMS2 Solution.

	Dimension 1.	Dimension 2.
1	.473	.550
2	.507	.448
3	.396	.727
4	.544	.361
5	.639	.363
6	.422	.394
7	.444	.534
8	.612	.507
9	.701	.316
10	.645	.237
11	.521	.427

TABLE 5.3: Subject space dimensional weights for the SIMS3 Solution.

	Dimension 1.	Dimension 2.
1	.720	.398
2	.231	.526
3	.651	.250
4	.629	.232
5	.696	.206
6	.611	.303
7	.727	.264
8	.573	.391
9	.703	.263
10	.475	.532
11	.719	.257

mar ked horseshoe effect. The high correlations of the complexity measure with both dimensions of that configuration suggest that the horseshoe itself may correspond to the complexity measure. Tables 5.1 to 5.3 indicate that there is quite a large amount of individual variation in the fitted dimensional saliences. The salience of dimension two is generally low for all three of the stimulus subsets.

Interpreting the INDSCAL Dimensions

In the preceding chapter it was found that the two-dimensional INDSCAL group spaces for SIMS1 and SIMWR could in general be adequately explained by one or two physical features of the Walsh stimuli. A method for interpreting the dimensions derived in an MDS analysis, in terms of a single feature for each dimension, has been outlined by Kruskal and Wish (1978, p.35 ff). It involves the use of Multiple regression with (in this case) the three INDSCAL dimensions as predictors and the physical feature as the criterion. The multiple correlation can be used as an estimate of how well the INDSCAL dimensions relate to the physical feature. The beta weights solved for in the multiple regression analysis can be converted to directional cosines (c_i) using the formula:

$$c_i = b_i \cdot \frac{1}{(b_1^2 + b_2^2 + b_3^2)^{½}}$$

where b_i is the i^{th} beta weight.

The direction cosines may then be used to represent the physical features as vectors in the INDSCAL group space. The analysis outlined above was performed on the experimental results and the direction cosines and multiple correlations before each feature and all three dimensions in each of the 3-D INDSCAL spaces are given in tables 5.4 to 5.6.

Two main points were taken into account in assessing the importance of a feature from an inspection of the direction cosines which relate it to an INDSCAL solution:

(1) The squared multiple correlation between the feature and the INDSCAL solution should be greater than .8.

(2) The feature should appear to relate to more than one of the INDSCAL solutions.

In addition, column sequency was included, although it was strongly related to only the SIMS1 solution, because of its theoretical importance as a basis for constructing the Walsh stimuli.

Tables 5.4 to 5.6 give an indication of the relationship between the INDSCAL solutions and the 34 features of the Walsh stimuli used in this thesis.

Features	Direction Cosines Dim 1	Dim 2	Dim 3	R^2
1	-.995	.019	.097	.972*
2	-.424	-.009	-.905	.861*
3	.351	-.600	-.719	.177
4	-.947	.222	.230	.766*
5	-.881	-.116	-.458	.813*
6	.839	.504	.204	.970*
7	-.588	-.809	.008	.273
8	.074	-.995	.071	.567
9	-.519	-.543	.661	.690*
10	-.896	-.071	.439	.676*
11	.024	-.703	.711	.601
12	.313	.137	.940	.849*
13	-.609	-.786	.102	.727*
14	.365	.106	-.925	.695*
15	.703	.711	.037	.594
16	.517	.836	-.184	.186
17	-.487	-.841	-.234	.439
18	.200	.872	.447	.062
19	-.217	.946	-.239	.053
20	.085	.913	-.398	.515
21	-.536	-.830	-.156	.809*
22	.805	.230	.547	.183
23	.489	.841	-.231	.653*
24	-.754	-.627	.196	.932*
25	-.880	.186	-.438	.878*
26	.489	-.328	.808	.370
27	-.978	-.047	-.202	.940*
28	-.982	-.009	-.186	.970*
29	-.762	-.058	.645	.485
30	.993	.117	.011	.918*
31	.679	-.727	-.101	.071
32	.090	.836	-.541	.289
33	-.610	.463	-.643	.784*
34	-.994	.087	-.064	.951*

Table 5.4 Results for the SIMS1 INDSCAL solution dimensions regressed on each of the 34 features. $^*F_{.01}(3,11) = 6.22$. The significance of the regression was assessed using an F test (with 3 and 11 degrees of freedom) and was found to be significant ($p < .01$).

Features	Direction Cosines Dim 1	Dim 2	Dim 3	R^2
1	-.291	-.811	.508	.598
2	.154	-.798	-.582	.922*
3	-.788	.192	-.585	.168
4	-.890	-.204	-.407	.499
5	.474	-.880	-.025	.721*
6	.400	.896	-.194	.933*
7	.247	.073	.966	.421
8	.077	.388	.919	.345
9	.255	-.454	.854	.375
10	-.945	.276	.175	.790*
11	.520	-.143	.842	.421
12	-.271	.822	.501	.754*
13	.110	-.067	.992	.578
14	.280	-.418	-.864	.724*
15	.842	-.539	.020	.480
16	-.491	.264	-.830	.348
17	.508	.174	.843	.131
18	.236	-.792	.563	.158
19	.759	-.282	.587	.559
20	-.836	.223	.501	.189
21	.096	-.017	.995	.465
22	.401	-.365	-.840	.467
23	.853	-.514	-.091	.452
24	-.218	-.056	.974	.896*
25	-.549	-.469	-.692	.696*
26	-.609	.787	-.101	.550
27	-.469	-.852	.232	.769*
28	-.529	-.847	.047	.907*
29	-.899	.194	-.393	.518
30	.550	.822	.145	.774*
31	.711	-.197	.675	.335
32	-.445	-.040	-.895	.617*
33	.392	-.521	-.758	.805*
34	-.553	-.709	-.438	.775*

Table 5.5 Results for the SIMS2 INDSCAL solution dimensions regressed on each of the 34 features. *These regressions were significant ($R^2 > 0$).

Features	Direction Cosines Dim 1	Dim 2	Dim 3	R^2
1	.977	.203	-.066	.671*
2	.934	-.358	-.021	.824*
3	-.564	.825	-.041	.068
4	.699	.713	.053	.715*
5	.767	-.627	-.135	.650
6	-.844	-.120	.523	.965*
7	.263	-.769	-.582	.615
8	.260	-.551	-.793	.212
9	.022	.149	-.989	.394
10	-.022	.948	-.319	.823*
11	-.510	-.037	-.859	.451
12	-.821	.539	.188	.819*
13	-.040	-.353	-.935	.237
14	.886	-.364	.288	.435
15	.390	-.741	.547	.420
16	-.067	.809	.584	.215
17	.869	.059	-.492	.099
18	-.067	-.917	-.393	.405
19	-.062	-.751	-.658	.276
20	-.104	.067	.995	.218
21	.115	-.481	-.869	.078
22	.668	-.719	-.194	.121
23	.281	-.859	.429	.391
24	.212	.339	-.917	.415
25	.931	.363	.034	.795*
26	-.602	.797	.053	.604
27	.985	.139	-.102	.767*
28	.963	.145	-.227	.928*
29	-.165	.937	-.308	.398
30	-.927	-.304	.220	.758*
31	.936	-.297	-.188	.055
32	.420	.452	.787	.113
33	.920	-.261	.292	.524
34	.935	.354	-.019	.870*

Table 5.6 Results for the SIMS3 INDSCAL solution dimensions regressed on each of the 33 features. *These regressions were significant (R^2 .0).

According to the criteria outlined above, the following six features should best be able to account for the similarity judgements made in SIMS1, SIMS2 and SIMS3:

1. column sequency
2. row sequency
3. squareness
4. average grain
5. component one
6. preference scale one.

Figures 5.4 to 5.6 depict the features which are most closely related to the INDSCAL solutions as vectors[1] inside the solution spaces for SIMS1, SIMS2, and SIMS3 respectively.

It can be seen that:

(a) Not all the INDSCAL dimensions (across the three solutions) can be interpreted in terms of the features used here.

(b) Some of the features are strongly, but nonetheless obliquely, related to the INDSCAL dimensions.

(c) The interpretation of the INDSCAL solutions in terms of the feature vectors changes across the three different sets of 15 stimuli which were used.

The informa tion in figures 5.4 to 5.6 is supplemented by Table 5.7 which relates eight of the features to each of the three INDSCAL solutions.

	SIMS1	SIMS2	SIMS3
Column Sequency	Dimension 1 $R^2 = .972$		Dimension 1 $R^2 = .671$
Row Sequency	Dimension 3 $R^2 = .861$	Oblique (2-3) $R^2 = .922$	Dimension 1 $R^2 = .834$
Squareness	Oblique (1-2) $R^2 = .970$	Dimension 2 $R^2 = .933$	Oblique (1-3) $R^2 = .965$
Average grain	Dimension 1 $R^2 = .676$	Dimension 1 $R^2 = .790$	Dimension 2 $R^2 = .823$
Component One	Oblique (1-2) $R^2 = .932$	Dimension 3 $R^2 = .896$	
Preference Scale One	Dimension 1 $R^2 = .970$	Dimension 2 $R^2 = .907$	Dimension 1 $R^2 = .928$
Complexity	Dimension 1 $R^2 = .940$	Dimension 2 $R^2 = .769$	Dimension 1 $R^2 = .767$
2-D Sequency	Dimension 1 $R^2 = .951$	Oblique (1-2-3) $R^2 = .775$	Dimension 1 $R^2 = .870$

Table 5.7 Interpretation of the SIMS1, SIMS2, and SIMS3 3-D INDSCAL solutions in terms of eight of the Walsh features.

Complexity and two-dimensional (2-D) sequency were added to the other six features because of their theoretical importance (cf. the discussion of the relationship of complexity and 2-D sequency to preference in chapter three). Preference scale one is strongly related to one of the dimensions in all three INDSCAL solutions. Complexity and 2-D sequency show a similar

1. The orientation of these vectors is given as $\cos^{-1}$ of the direction cosines. Following Kruskal and Wish (1978), the present method consists of selecting the dimension on which the feature has the largest direction cosine (providing that that feature has a squared multiple correlation of at least .8) and then positioning it according to the angle implied by that cosine. Collapsing the vector into the 2-D plane in this fashion means that the angle between the vector and the second dimension (the one with the smaller of the two direction cosines should be consulted to check the actual size of the direction cosine for a more accurate measure of where the feature vector is located with respect to the second (further away) dimension. (The discrepancies in figures 5.4 to 5.6 are generally a matter of two to three degrees. Larger discrepancies have been marked with asterisks).

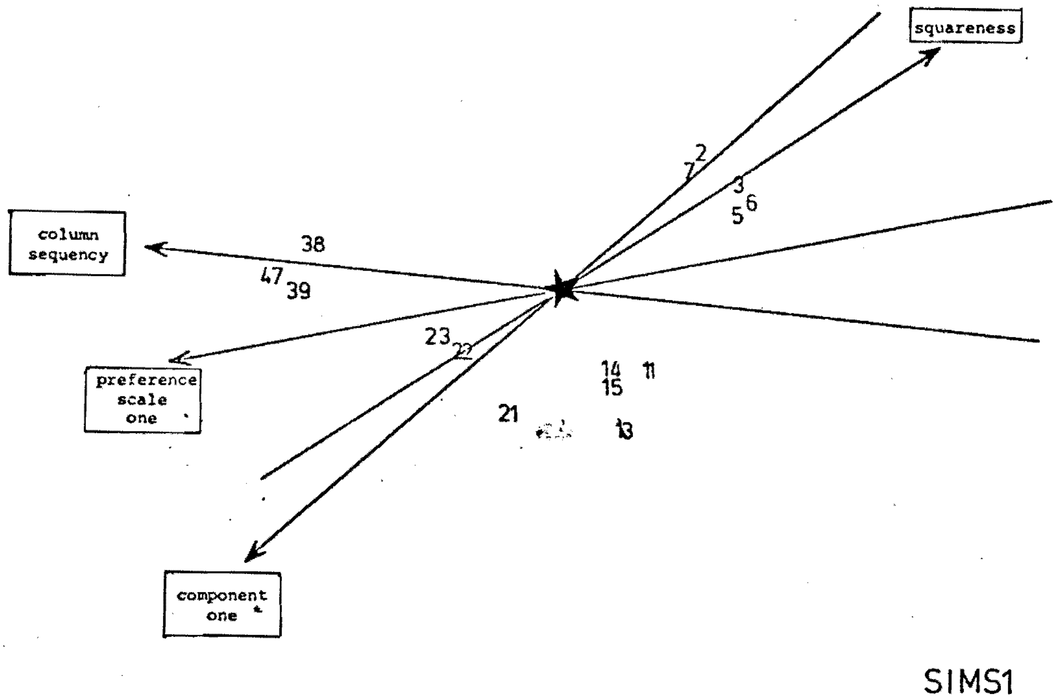


Figure 5.4(a) four of the feature vectors in the dimensions one (x-axis) and two of the SimSl INDSCAL solution.

*component one is oriented with respect to diml.

It is at an angle of 51° w.r. to Dim2

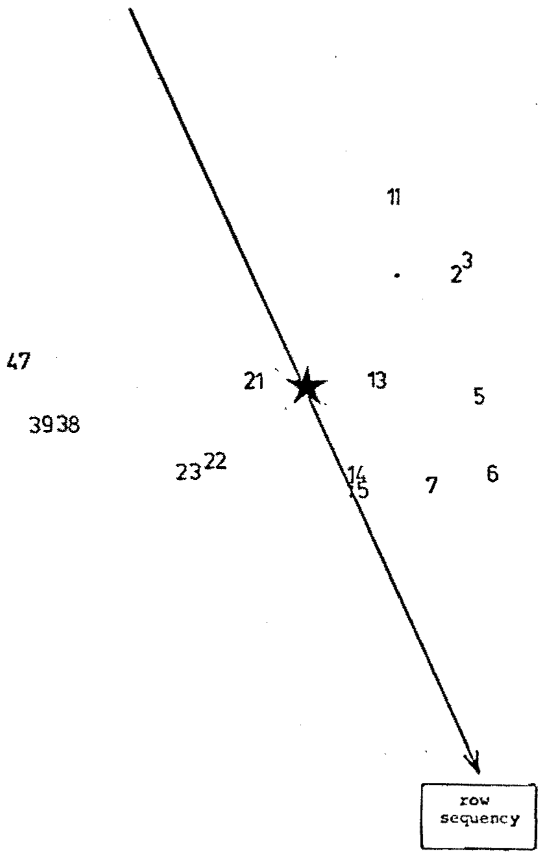


Figure 5.4(b) dimensions 1 (x-axis) and *3* of the SimSl solution

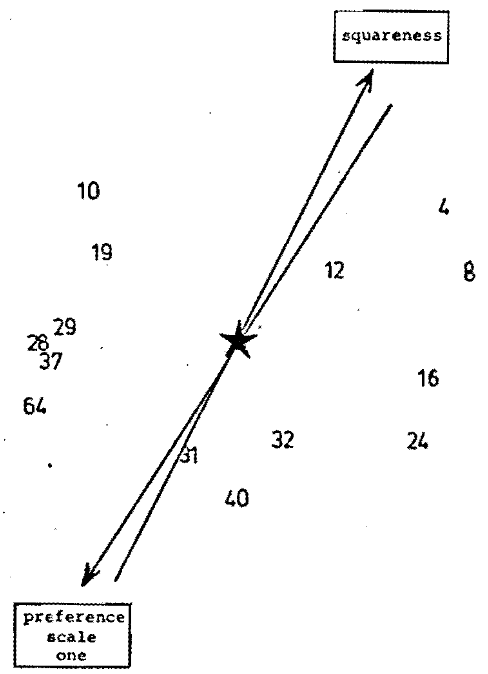


Figure 5.5(a) two feature vectors located in dims1 (x-axis) and two of the SimS2 solution

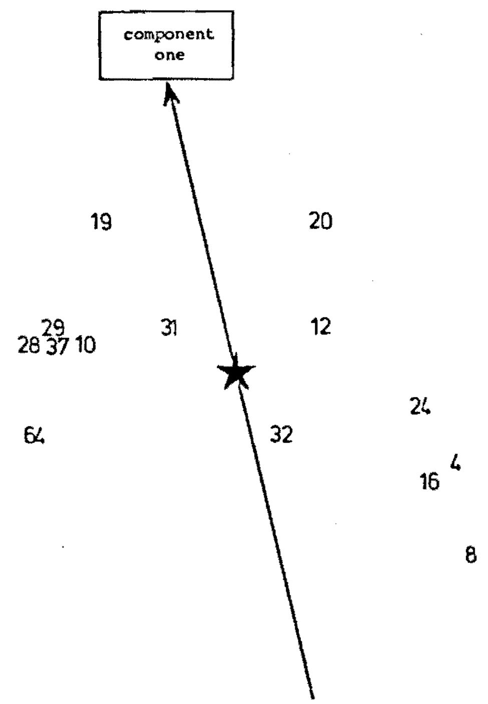


Figure 5.5(b) component one located on dim 1 (x-axis) and three of the SimS2 solution

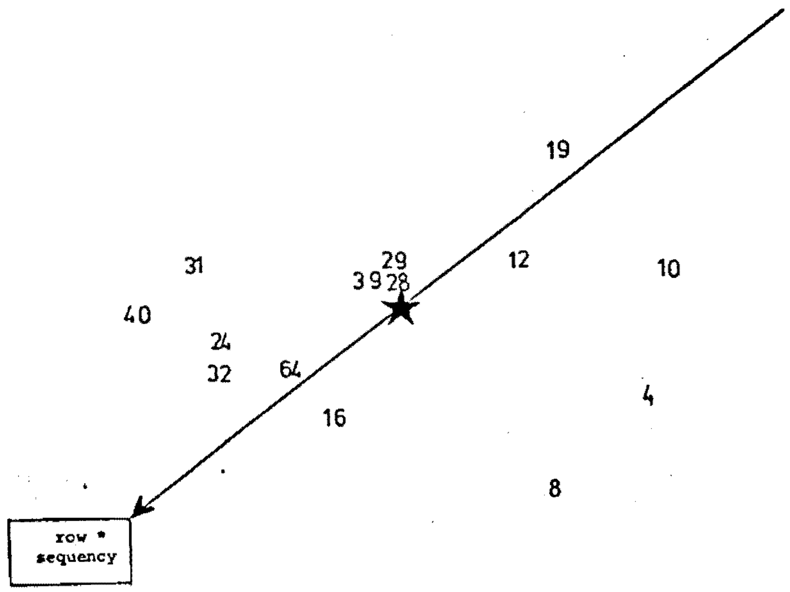


Figure 5.5(c) row sequency in dims2 (x-axis) and three of the SimS2 INDSCAL solution

*row sequency is at an angle of 54° to dimension three

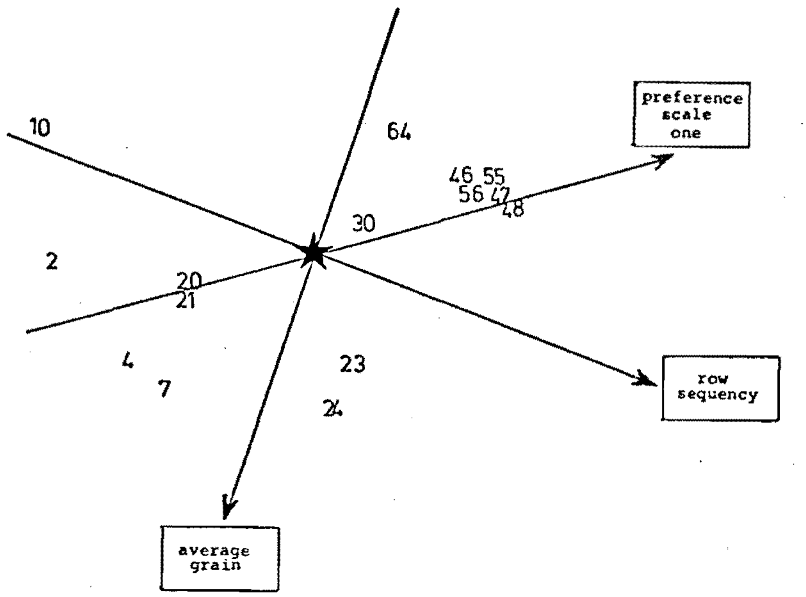


Figure 5.6(a) three features located in dims1 (x-axis) and two of the SimS3 solution

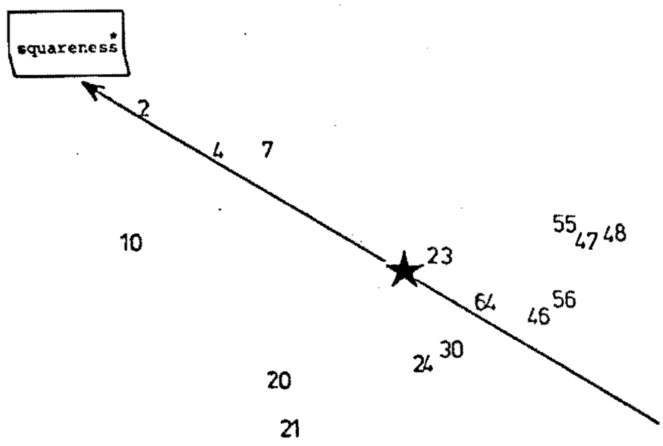


Figure 5.6(b) squareness (feature 6) located in dims1 (x-axis) and three of the SimS3 solution

* squareness is at an angle of 50° with respect to dim3

(but slightly weaker overall) pattern of relationships across the three solutions. The effect of column sequency is also smaller than, and accounted for by, the effect of preference scale one. Component one is derived from statistical relationships between the features and would not normally be expected to be an underlying feature in similarity judgement. Squareness and average grain both have relationships with the INDSCAL solutions which cannot be completely accounted for in terms of row sequency or preference scale one.

The apparent predominance of preference scale one as an underlying feature in similarity judgements may in fact be largely a result of its relationship with complexity and two-dimensional sequency, rather than being due to the effect of preference on similarity judgements. One way to check for this possibility is to see if differences in a person's preference scale have a systematic effect on similarity judgements. Although some of the features derived in chapter two are reflected in the dimensions of the INDSCAL solution, there appears to be a certain amount of disagreement between the feature representation of the Walsh stimuli developed in chapters two and three of this thesis and the features indicated by the INDSCAL solutions.

In many cases the 34 Walsh features considered here are either obliquely related to the INDSCAL dimensions or else they are largely unrelated to the INDSCAL solution. Previous studies (e.g. Wish and Carroll, 1974) have found INDSCAL dimensions which correspond closely to physical features of the stimuli used. This is not always the case in the present study, although *some* of the INDSCAL dimensions are directly interpretable. This suggests that the feature set used here may be deficient in one or more crucial features, and that, in some cases, the INDSCAL method, while identifying the main features underlying the similarity judgements, has nevertheless rotated them. While the orientation of axes is unique in the INDSCAL model it may be the case that the application of the INDSCAL model where the features are significantly correlated, and where different participants may base their judgements on difference features, may lead to an apparent rotation of the features underlying the judgements. This latter possibility would have important implications for the metatheory of three-way scaling as the unique orientation of the INDSCAL solution has been an important argument for using it in preference to other three-way MDS methods. The alternative explanation for the pattern of results for SIMS1, SIMS2, and SIMS3 (as analysed by INDSCAL) is that in some cases (for example, SIMS1 dimensions one and three; SIMS2 dimensions two and three; SIMS3 dimensions one and two) the participants base their judgements directly on the physical features while the remainder of the features used are secondarily derived and only obliquely related to physical features such as those used here.

Rösler (1979) has recently questioned the usefulness of the INDSCAL method in identifying individual differences. He suggested that a least squares method such as INDSCAL may have no more psychological meaning than a corresponding set of principal components. On the whole such a suggestion is probably incorrect (note for instance the featural interpretations of the INDSCAL solutions given above) but a systematic simulation study of the CANDECOMP algorithm (a key part of the INDSCAL method) does appear to be necessary, as noted by Rösler. On the other hand, Rösler's (1979 , pp 165-166) remarks about the relative usefulness of similarities and rating scalings are in direct conflict with the present orientation (note particularly the failure of E1 in chapter two to produce meaningful results and the central role ascribed to similarity in the present cognitive framework of perceptual learning). This apparent conflict may be at least partly due to the use of conceptual as against perceptual stimuli. Rating scales are themselves concepts and may well be more appropriate with stimuli that are overlearned and largely culturally determined (cf. the comments on the possibility of a unique representational structure made at the beginning of this chapter).

The next step in the present analysis would be to use three-mode scaling (Tucker, 1972) to assess individual differences in similarity judgement with a model that allows the oblique rotation of axes. This whole issue of the extent to which the advantages of unique axis orientation outweigh the problems associated with modelling psychologically distinguishable, but nevertheless correlated, features with orthogonal dimensions deserves serious attention, although it is beyond the scope of this present thesis.

Clustering Analysis

Average linkage hierarchical cluster analysis (Sokal and Sneath, 1963) was run on the results of one of the participants from each of the SIMS1, SIMS2, and SIMS3 experiments using BMDPIM (Dixon, 1975). Figure 5.7 shows the clustering solution for the participant with the highest correlation (r = .875) between observed and computed scores embedded within the two-dimensional INDSCAL solution for SIMS1.

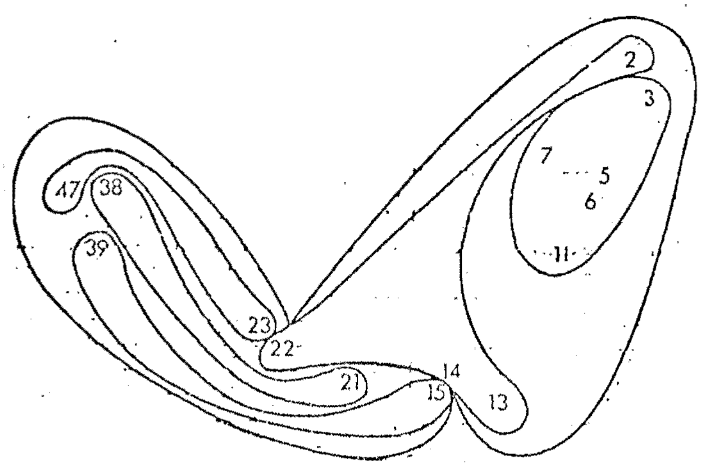


Figure 5.7 Clustering solution from the best fitting SIMS1 participant embedded in the SIMS1 2-D INDSCAL solution.

It can be seen that the similarities, as implied by the clustering solution, appear to be somewhat distorted by the INDSCAL solution. This participant's dimensional saliences were .851 and .120 for dimensions one and two respectively. Figure 5.8 shows the same clustering solution embedded in the transformed group space where the general incompatability between the INDSCAL and clustering solutions is also in evidence.

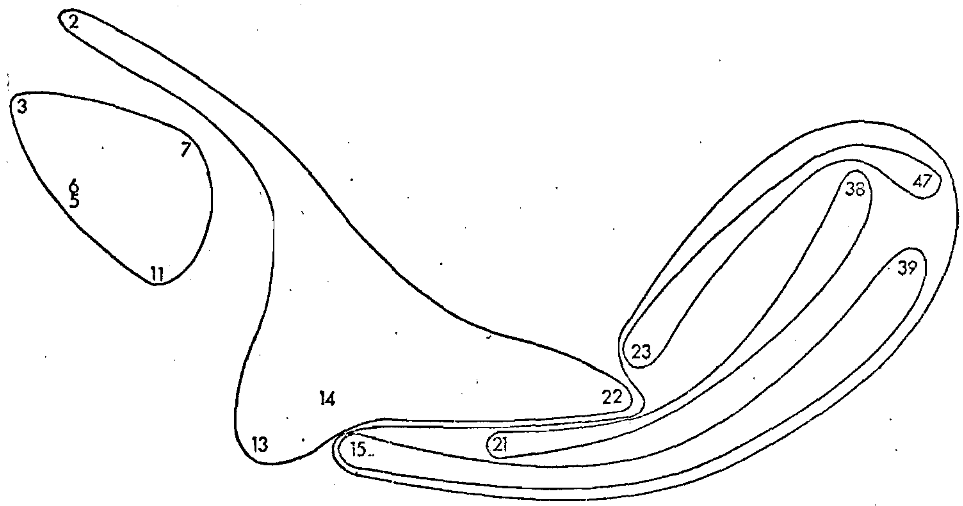


Figure 5.8 Clustering solution for the best fit SIMS1 participant embedded in the transformed INDSCAL (the reversal of the x-axis should be ignored).

The participant chosen to represent SIMS2 had both a high correlation (r = .796) and relatively large saliences (.612 and .507 for dimensions one and two respectively). Figure 5.9 shows the clustering results embedded in the two-dimensional (2-D) INDSCAL space for SIMS2.

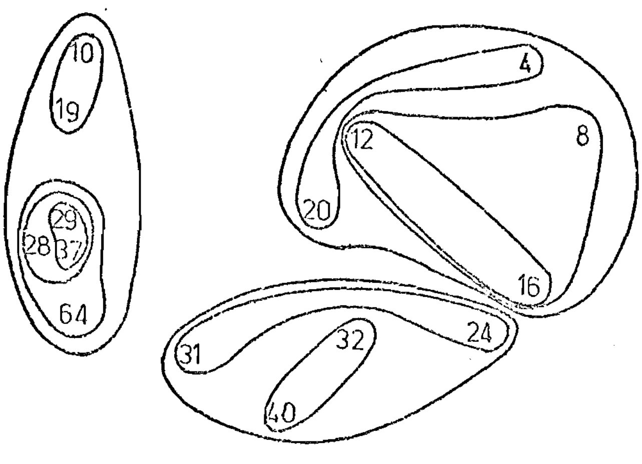


Figure 5.9 Configurational representation of the SIMS2 results.

The SIMS3 representative also had good fit (r = .857) and relatively high dimensional saliences. Figure 5.10 shows the SIMS3 INDSCAL 2-D group space with his clustering solution embedded in it.

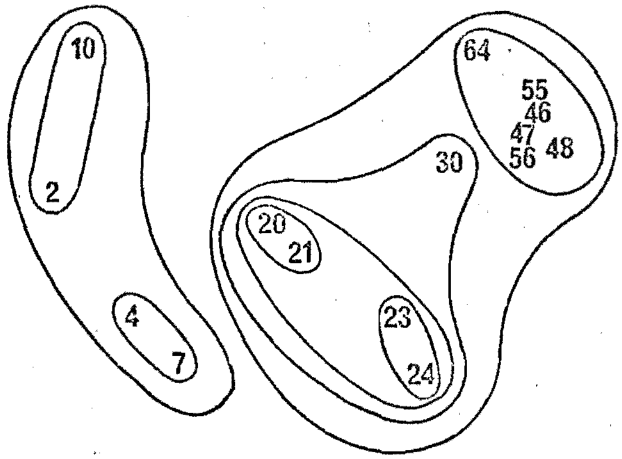


Figure 5.10 **Configurational representation of the SIMS3 results.**

The apparent incompatability of the clustering and INDSCAL solutions for SIMS1 may be due to the low salience of dimension two for the person whose results were clustered. It is surprising however that the INDSCAL solution appeared to give a good fit to that person's result in spite of the apparent distortion of the similarities data (as represented by the clustering solution) in the INDSCAL solution. The point will not be pursued here as it is somewhat peripheral to my thesis, but it might be worth investigating whether low saliences for an individual may indicate that the INDSCAL model is not appropriate for that person even in cases where the fit of model predictions to the data appears to be quite good.

Summary

Similarity judgements are a convenient method of deriving the configurational representation of a set of stimuli. This should not obscure the fact that similarity judgement is a decision process which may prove useful in testing and refining perceptual-cognitive theories.

The INDSCAL analyses of the SIMS1, SIMS2, and SIMS3 results generally reproduced (though sometimes in rotated form) one or more out of a subset of eight of the features used in this thesis to quantify the Walsh stimuli (as shown in Table 5.7). Some of the INDSCAL solution dimensions did not appear interpretable, and these dimensions may have resulted from violations of the INDSCAL model in the experimental results.

CHAPTER VI

The present chapter is an extension of the theory and experimentation outlined in chapters four and five. Chapters four, five and six, taken together, represent an attempt:

1. to generate a more specific and predictive account of the effect of task demand effects in perceptual — cognitive processing,
2. to develop experimental paradigms which would allow the detailed investigation of task demand effects,
3. to quantify the differential effect of four different paradigms (using the Walsh stimuli) on responding, where the paradigms were designed to differ in the pattern of processing strategies that they each elicited in the participants.

Chapter four quantified the effects of selective attention on the perception of the Walsh stimuli. Both the INDSCAL model and the regression model were found to be sensitive to the effect of the selective attention instruction.

Chapter five developed configurational representations of the three subsets of Walsh stimuli as well as interpreting similarity judgements as part of the general cognitive framework being used in this thesis.

The theory and findings of chapters four and five may now be used to investigate what happens as the experimental task demands change. It was suggested in chapter four that changes in attentional strategy mediate the effect of task demands on the pattern of responding. Consequently, one point of interest in the present chapter will be the extent to which changes in the pattern of responses due to differing experimental task demands appear to be related to changes resulting from the selective attention instruction. The configurational representations and similarity models of chapter five may be used as a baseline for assessing the nature of the changes in perceived stimulus structure due to task demand effects. The first part of this chapter will consider a number of modifications which can be made to the basic type of similarity judgement elicited in chapter four. These modifications will be accounted for in terms of a wider cognitive framework which includes the matching tasks used by Posner and others.

Modifications in the Similarity Judgement Paradigm

Most similarity experimentation in the past has used the simultaneous presentation of stimulus pairs. There are a number of ways of altering the basic paradigm (simultaneous presentation for a relatively long period of time) which should be of interest in the investigation of judged similarities within a cognitive framework. For instance, the exposure time may be varied from tachistoscopic to indefinitely long durations. Dornic, Künnapas and Bratfisch (1970) conducted two experiments which estimated time effects on similarity.

In their first experiment, pairs of simple visual stimuli were exposed for five different time periods which ranged from one to twenty msecs. They found that similarity decreased with increasing exposure time. According to Dornic *et al.,* the very short exposure times prevented the participant from using rational strategies (such as counting the number of sides in the figure) in making the similarity judgement. In the second experiment run by Dornic *et al.* (1970) the participants compared the similarity of the first and last stimulus in a series of seven figures. The whole series was repeated five times. Similarity was found to decrease with repeated presentation of the series.

The results of Dornic *et al.* are only suggestive because of the following deficiencies in their study: only a small number of stimuli were used, there was a lack of variation in the stimulus series used in the second experiment and the results obtained on a 100-point rating scaling were averaged across participants.

The relevance of the work of Dornic *et al.* (1970) for the present thesis is that they distinguished between perceptual and cognitive similarity on the basis of variations in the experimental paradigm:

"Essentially different mechanisms are involved in Experiment II, where the estimates were limited by memory rather than by sensory or perceptual discriminative capacity."

—Dornic *et al.,*(1970, p.7)

Künnapas (1968) found that the subjective similarity between letters of the Swedish alphabet was the same when the letter pairs were presented visually as when their memorial representations were cued. Künnapas attributed this lack of difference to the overlearned nature of verbal material, but there may be auditory-visual confusions also involved. Simultaneous visual presentation of stimulus pairs and memorial comparisons can be regarded as the two extremes on a continuum of memory involvement in similarity judgements.

Quadratic Similarities

One set of possible modifications to the similarity judgement paradigm (briefly touched on in the previous section) affect what may be called the location of the stimuli. For instance, delayed presentation of the second stimulus changes the *temporal* location of the second stimulus, relative to the first. *Spatial* location may also be changed, although the amount of relative positioning of a subset of stimuli to be presented in a trial is constrained by the ability of the participant to scan the relevant stimuli during the time (or times) available.

The second set of modifications can be referred to as context. The subset of stimuli presented in a particular trial can change with respect to *numerical* context (that is, what is the total number of stimuli presented on that trial). There may also be changes in the *structural* context. Thus the stimulus subset may differ in terms of the categorical and dimensional structure of the stimulus. Once the inherent flexibility of the similarity judgement paradigm is realised, it becomes possible to develop a whole range of similarity judgement paradigms, each of which will have different storage and processing requirements.

Two forms of stimulus context (the total stimulus set and the subset of stimuli used in a given trial) were mentioned in chapter five. The context mentioned in the present section refers solely to the subset of stimuli used in a given trial.

Structural context has been extensively considered by experimental psychologists. An example of this is the research on the effect of separability and integrality of stimulus dimensions on similarity judgements (cf. chapter four). Numerical context, on the other hand, has not generally been considered. Psychological methods such as the method of triadic combinations (Richardson, 1938) and the complete method of triads (Torgerson, 1952) are generally used as a convenient method of generating pairwise similarities data.

Gregson's (1969, 1970) quadratic similarities paradigm is a similarity judgement method which incorporates modifications in terms of both spatial location and numerical context. One of the major problems in using a stimulus subset of size greater than two in a similarity judgement is that it becomes difficult to specify what the similarities are based on. There do not appear to be any studies which have used triad-wise (that is the mutual similarity of three objects taken together) similarity judgement, let alone a theory of what sort of comparison process such a judgement would elicit. Thus the available theories of similarity imply that similarity is a binary operation between two entities or combinations of entities. The binary nature of similarity judgement can be represented in the levels of Relative Judgement notation of Gregson (1975, pp.16-24).

Consider the situation where there are three stimuli in the stimulus subset. One data collection method might consist of requiring the participant to assess all three of the pairwise similarities. If there were n stimuli and all possible triads (${}^{n}c_{3}$) were represented, a particular pair of stimuli would appear in n - 2 contexts. This paradigm would be an interesting way of assessing the effect of context (within the presentation subset) or pairwise similarity judgement.

An alternative data collection strategy would be to nominate one of the stimuli as being of particular interest and then obtain the two pairwise similarity judgements between that stimulus and the other two stimuli in the subset. This introduces an inherent asymmetry into the judgement (cf. the section on asymmetry in chapter four) which can be captured by referring to the chosen stimulus as the referent, and the stimuli against which it is judged as the referees. Quadratic similarity judgements are made when there is a subset of four stimuli composed of three referees and one referent, the latter being chosen from a set of referents. Gregson (1969) made the additional modification of having the participant respond by placing a marker representing the referent within an equilateral triangle grid plan (with the three referees represented one at each of the vertices) so that its location represented the perceived relative similarities of the referent and the referees.

The quadratic similarities tasks require three simultaneous pairwise comparisons of the participants and they can be viewed as a modification of the basic similarities paradigm by increasing task demands to a point where the participant has to change his/her attentional strategy:

"The QS task with 4D stimuli is near to or above the information-handling capacity of many subjects when done as a timed task. Subjects can perform the required judgements in a suboptimal fashion by ignoring some of the presented informa tion; for example, they may cut out one subset or dimension, which is effectively giving it very low weighting and compensatorily give to the remaining subsets large weighting."

— Gregson (1975, p.173).

Gregson (1972) used the following change in the relative temporal location of the referees and referent. In one condition the referees and referent were all simultaneously presented for one minute.[1] In the second condition the referent only was presented for two seconds in the middle of the one-minute period for which the referees were exposed.

The Quadratic similarities tasks appear to represent an upper level of complexity in the modified similarity paradigms used in the previous literature. The incorporation of such a task within a cognitive framework is not at all obvious. The next section will look at a task that has become popular with cognitive psychologists since its formulation in 1967, and will be followed by an attempt at developing a general theoretical framework for cognitively-oriented similarity judgements.

The Posner Task

The effect of moderate levels of memory involvement on similarity judgements does not appear to have been systematically investigated. However, there is one paradigm (commonly called the Posner task) which has been used to investigate different levels of memory involvement in same-different judgements involving letter pairs. The paradigm is as follows:

A pair of letters is presented simultaneously. The participant is required to press one key if the stimuli have the same name and another if their names are different. If the two are physically identical (for example, AA), it is logically possible to make the match based upon the visual form. If the letters do not have the same physical form (for example Aa), the match must be based on a previously learned correspondence. Posner and Mitchell (1967) used this paradigm and found that the time taken to match physically identical

1. There was also a change in the brightness of the display after 30 seconds' exposure time but this change does not appear to have been a key feature of the paradigm. Brightness has, however, been shown to have an important effect of masking (Turvey, 1973) and may well have an effect in certain kinds of delayed similarity judgements.

letters was about 80 msecs faster than letters having only names in common. (As is generally typical of present cognitive research, it was the pooled results which were reported).

Posner and Keele (1967) found that this name versus physical code effect was lost if the second letter was presented after a two-second delay (instead of simultaneously with the first letter). Posner, Boies, Eichelman and Taylor (1969) modified the paradigm still further by including an interpolated arithmetic task. Subsequent studies based on this paradigm yielded contradictory findings, apparently owing to differences in experimental design. (Kroll and Parks, 1978).

Two overall conclusions can be made from the various studies using interpolated tasks in the visual matching of letters under delayed presentation:

1. The visual code resides in permanent memory (the system of categories in the present terminology) and maintenance activities are needed to "protect the accessibility" of this visual code (Kroll and Parks, 1978).
2. The maintenance activity required to protect the accessibility of the visual code is disrupted when the digits in the interpolated (arithmetic) task are presented visually, but not when they are presented auditorally (Proctor 1978).

These conclusions are based on same-different judgements using a restricted set of stimuli. Kirsner and Sang (1979) generalised the paradigm still further by using a variety of Letraset type fonts for the stimulus letters. Thus the physical code matches (AA) were between capital A in different type fonts. Varying levels of similarity between the type fonts were used in different trials of the experiment. Previous studies had found that when a single target letter was presented, the reaction time advantage for the physical match condition declined from approximately 90 msecs at an ISI (inter-stimulus interval) of .5 secs, to 10 msecs or less when the ISI was increased to 2 seconds. This phenomenon shall be referred to as *convergence* in the following discussion.

Kirsner and Sang's (1979) Experiment four was designed to assess the effect of ISI, similarity (between the stimulus letters), and 'context' on the phenomenon of convergence. Context was a factor in the design, with two levels. The first level consisted of a single task: either a physical *or* name match. The second level consisted of a dual task where the participant was required to give a name match followed by a physical match. The results of this experiment are expressed by the authors as follows:

"To summarize, although the results of Experiment 4 demonstrate that increased attention to the form of the target letter can reduce convergence between the physical match . . . and name match . . . conditions used here, the change is apparently due to reduced accessibility for the name code in the long ISI condition rather than to increased accessibility for the visual code . . . (A)lthough the visual and name codes may be competitive when the question of accessibility is considered . . ., there is no suggestion that the two codes compete for a limited retention capacity per se."

— Kirsner and Sang (1979, p.272).

Two general conclusions (based on all four experiments run by Kirsner and Sang) are relevant to the present discussion:

1. The degree of similarity of the target and probe letters in the Posner task has systematic effects on speed and accuracy.
2. Convergence results from the combination of a non-optional decrease in the accessibility of the visual code and an optional increase in the accessibility of the name code. The terms optional and non-optional refer to whether or not accessibility may be modified by controlled processing (Shiffrin and Schneider, 1977).

The following sections will develop a theory of delayed similarity judgements where such judgements are viewed as variants on the generalised similarity paradigm of which the Posner task is also a special case.

The Posner task viewed as a modified similarities paradigm

The work of Kroll and Parks (1978) and Proctor (1978) on the one hand, and Kirsner and Sang (1979) on the other, represent two somewhat different approaches to the use of Posner tasks in the study of visual short term memory. Within the present framework, these approaches both study the memory schemata which are activated by visually presented stimuli. The apparent decrease in the accessibility of the visual code would result from the transience of the activated memory schemata whereas the optional increase in the accessibility of the name code may arise when the stimulus event is incorporated into the representational structure.

One apparent conflict between Kroll and Parks (1978) and Kirsner and Sang (1979) concerns the optionality of the decrease in the accessibility of the visual code. The view taken here is that the activated memory schemata decay or 'de-activate' monotonically over time but that the rate of decay may be changed according to alterations in the experimental paradigm.

A generalised similarities paradigm can be envisaged where variation in the experimental task will depend on the modifications made to six parameters in a similarities task. Five of these parameters are spatial location, temporal location, numerical context, structural context (all mentioned previously in the section on quadratic similarities) and the method of response. The method of response is the response actually made by the participant, and could, for instance, be a rating on a seven-point rating scale. Incidental measures such as reaction time and correctness may be inferred by the experimenter. The sixth parameter is the set of instructions given to the participant.

As an example of instructional set, the participant may be required to be as fast *and* as accurate as possible, or (s)he may be asked to attend selectively to certain aspects of the stimulus. In addition to the six parameter paradigm outlined here, it is also possible to include an interpolated task such as the mental arithmetic required by Posner et al (1969). It is possible to incorporate the Posner task as a modified similarity judgement with a particular pattern of the six task parameters outlined above. The value of these parameters for the Posner task will now be made explicit:

(1) *numerical context* = 2 for the basic paradigm (Kroll and Parks, 1978, used a numerical context of four).

(2) *structural context.* (Sc). In some cases the stimulus structure is *explicit* in that the participants are told to respond to category differences between the stimuli (for instance, vowel versus consonant). In other cases stimulus structure may be *implicit* as in the Kirsner and Sang (1979) study where the stimuli were designed to have several levels of similarity within each category . Given that there are C1 explicit categories of stimuli, a total of C1 numbers may be used to characterise the stimulus structure in the present formulation. For instance

Sc = (2,2,2)

would indicate three categories of stimuli with two levels of similarity within each category.

(3) *temporal locations.* The ISI is of course a standard parameter in Posner tasks. No distinction appears to have been previously made, however, between simultaneous presentation of the stimulus pair (ISI = ϕ in the present notation) and sequential presentation where the second stimulus is presented as soon as the first stimulus is removed (ISI = 0).

If there were more than two stimulus presentations in a trial, then there would not be just one ISI but several. In such a case the relative temporal location of each stimulus can be indicated as the time interval (or intervals) during which it is presented relative to the time of the first stimulus presentation (i.e., the start of the trial). This notion of temporal location is illustrated schematically in figure 6.1 for the cases of the basic Posner task and one of the Quadratic Similarities tasks used by Gregson (1972). Figure 6.1 illustrates the manner in which the present general framework should be able to reveal fundamental equivalences in experimental paradigms which have previously been regarded as quite distinct.

(4) *spatial location.* Previous studies have not used variations in spatial location as part of the design of Posner tasks. Generally speaking, side by side orientations are used for the simultaneous presentation of stimulus pairs.

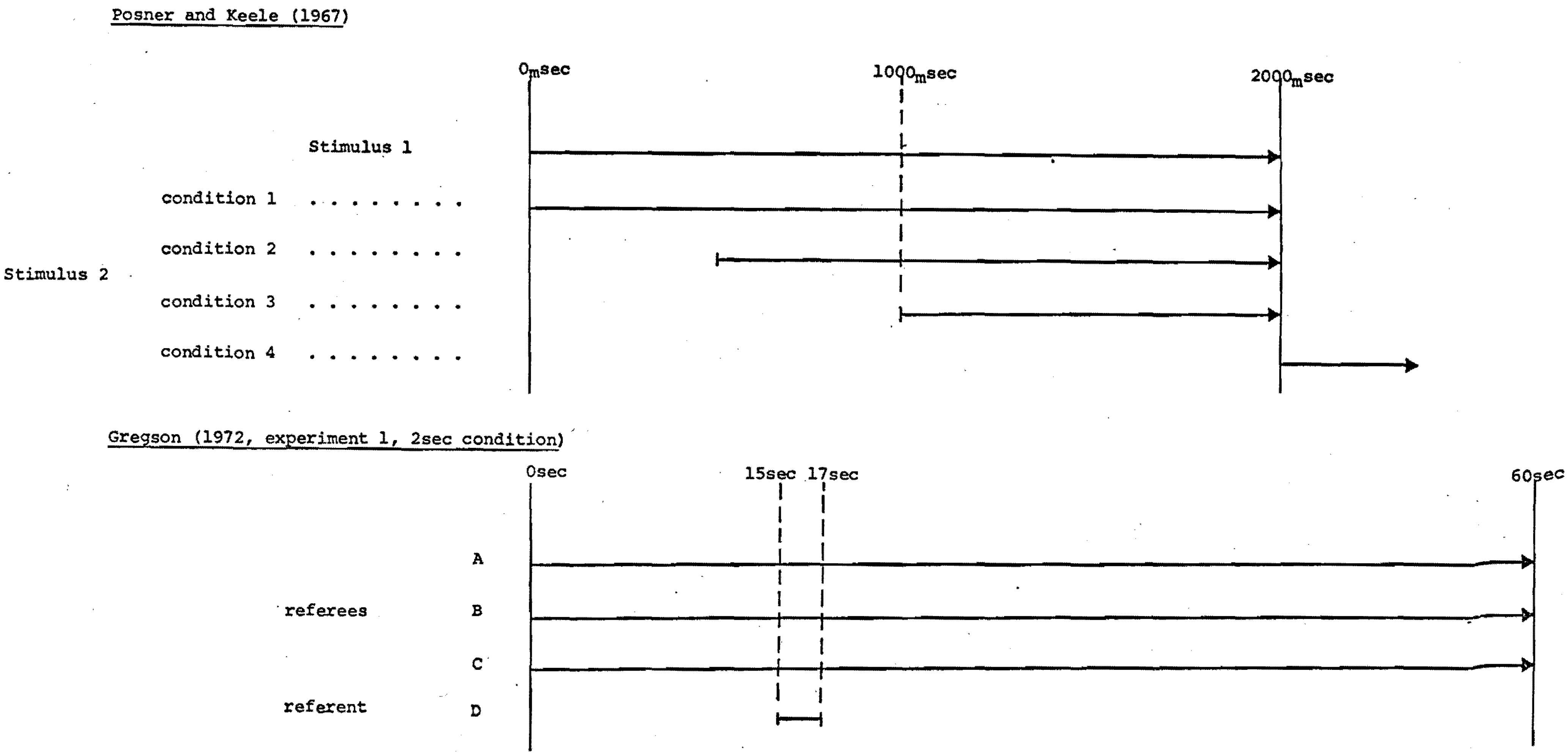


Figure 6.1 A representation of the relative temporal location of the stimuli in a single trial of two different experimental paradigms.

(5) *instruction set.* In most Posner tasks both speed and accuracy are emphasised.

(6) *the method of response.* The response consists of a same-different judgement.

Same-different judgements and similarities

The chief difficulty in incorporating Posner tasks within a similarities framework is the need to account theoretically for the relationship between same-different and similarity judgements. Gregson (1975) considered the relationship between same-different and similarity judgements. He suggested a certain compatability between the perceptual-cognitive models required to explain Posner tasks and these models need to account for similarity generation.

One way of formalising this compatibility is to posit the existence of a general similarities continuum which includes same–different judgements as a special degenerate case where the information is reduced to one bit.

A same-different judgement may then be seen as an enforced dichotimisation of that continuum into two fuzzy sets (Zadeh, 1965). The fuzzy boundary for the dichotomy could be mapped one-one onto some closed segment of the similarity continium. The next section will introduce a new form of modified similarity judgement which will be developed as a special case of the generalised similarity task (with its six parameters of variation) described in this section.

Delayed Similarity Judgements

A delayed similarity judgement as used in this thesis is a special form of generalised similarity task. The values of the six task parameters for the experiments reported later in this chapter are given below.

(1) numerical context = 2.

(2) structural context. Arguably the most obvious structure relates to the row and column sequencies (vertical and horizontal stripes) of the fifteen Walsh stimuli used in each experiment. There is no explicit relationship between the stimulus categories in the stimulus subset (two stimuli per trial) and the response categories, as there is in the Posner task where responses are either correct or incorrect. If Walsh sequency structure is not assumed then the structural context can be expressed as:

$$Sc = (1,1,1,1,1,1,1,1,1,1,1,1,1,1,1)$$

where the stimuli are not assumed to have any similarity/category structure in advance of the experiment.

(3) temporal location. The temporal locations of the two stimuli are represented in figure 6.2, for the SIMS1, SIMD1, and SIMD8 experiments.

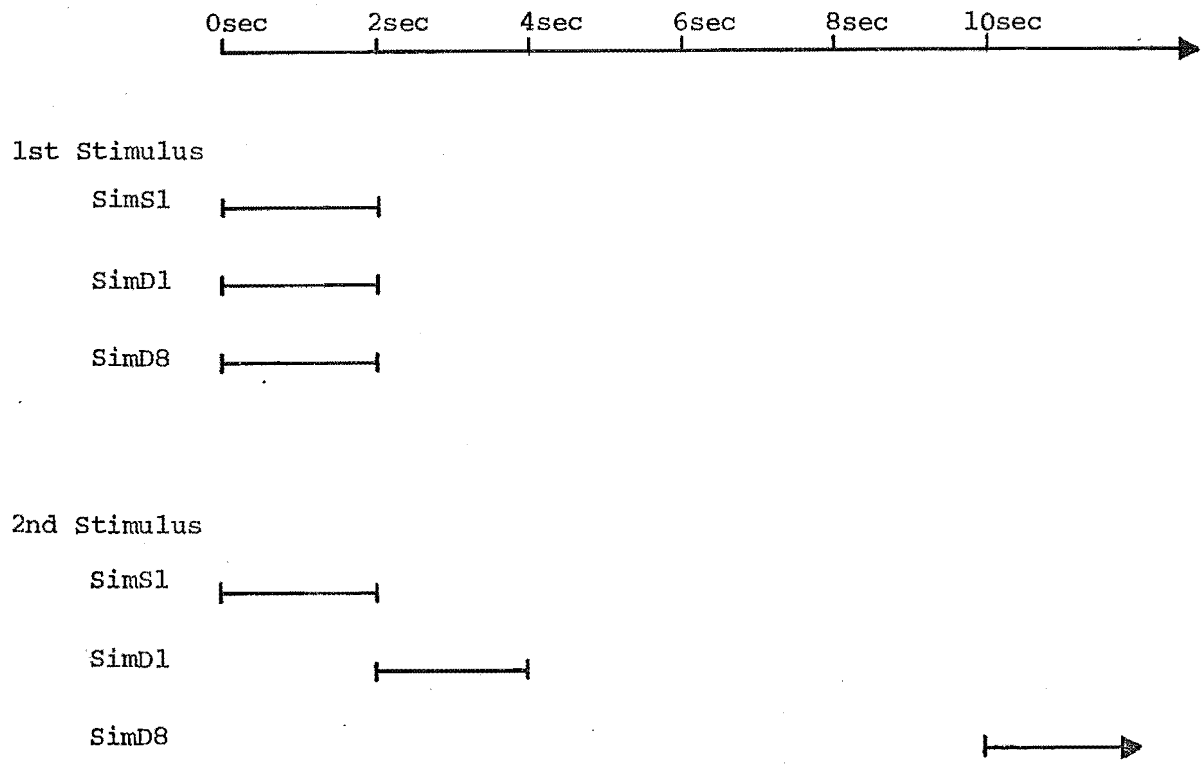


Figure 6.2 Relative temporal location of stimuli in single trials of the SIMS1, SIMD1, and SIMD8 experiments.

(SIMS1, SIMS2, and SIMS3 had the same relative temporal locations, while SIMD1, SIMD2, and SIMD3, had the same relative temporal locations as each other — but different from the relative temporal locations of SIMS1, SIMS2, SIMS3).

(4) spatial location. The stimulus pair is presented side by side regardless of delay between the first and second stimulus presentations.

(5) instruction set. No additional instructions other than an explanation of the method of response (see below).

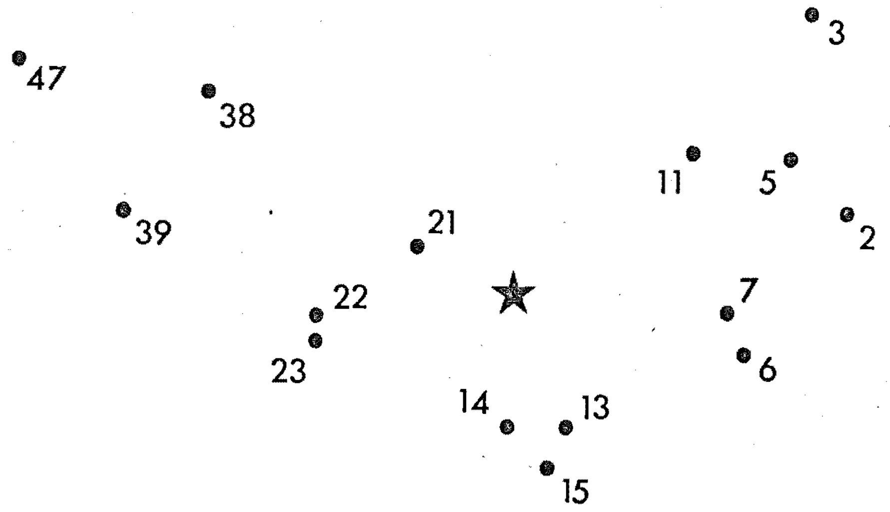


Figure 6.3 Two-dimensional INDSCAL solution for the SIMD1 results.

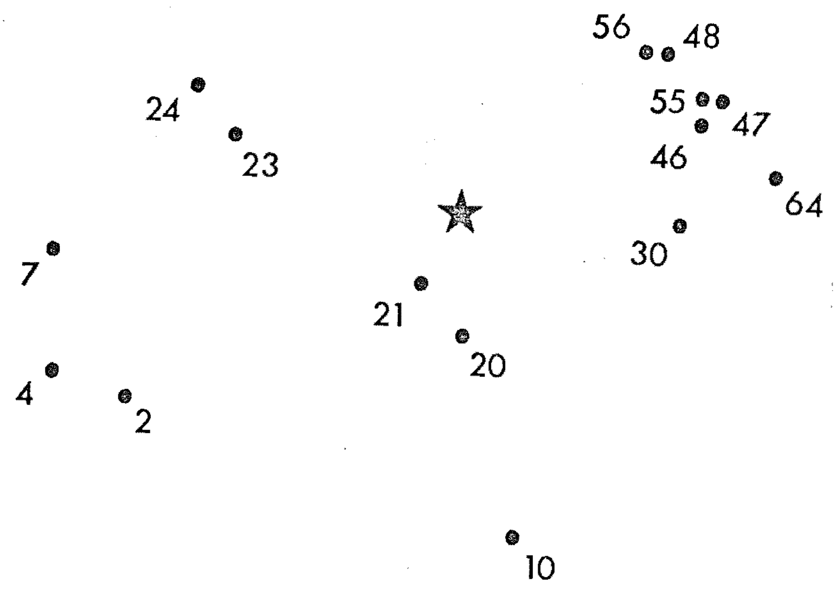


Figure 6.4 Two-dimensional INDSCAL solution for the SIMD3 results.

(6) the method of response. Rating on a seven point scale with end labels as described in chapter four (completely different, completely similar).

Experiments SIMD1, SIMD2, and SIMD3.

12 participants were used in SIMD1, eight in SIMD2 and six in SIMD3. The three experiments had the same design, and stimulus material, as the corresponding experiments (SIMS1, SIMS2, and SIMS3) which used simultaneous presentation, except that the second slide was not shown until the first slide had been closed off (ISI = 0). The duration of presentation for each of the slides was approximately two seconds. SIMS1 and SIMD1 used the same set of participants with replication order approximately balanced across the two experiments.

Results

The data obtained from each of the three experiments were analysed using INDSCAL. Two and three dimensional solutions were obtained. Figures 6.3 and 6.4 show the two-dimensional INDSCAL group space solutions derived from the SIMD1 results. The mean square correlation coefficients (a measure of average fit computed by INDSCAL) were low for the SIMD1 (.222) and SIMD2 (.137) results in comparison with the coefficient for SIMD3 (.578). The bad fit of the INDSCAL solution to the SIMD2 results may have been due to an intermittent fault on one of the projectors which was not discovered until after SIMD2 was run. Consequently only the SIMD1 and SIMD3 results will be considered further.

Interpreting the INDSCAL dimensions.

Multiple regression was used to represent each of the Walsh features in terms of the three dimensional INDSCAL solution spaces of SIMD1 and SIMD3, as was done previously for SIMS1, SIMS2, and SIMS3. The direction cosines and multiple correlations between each feature and all three dimensions in the SIMD1 and SIMD3 INDSCAL solution spaces are shown in Tables 6.1 and 6.2.[1] Table 6.1 shows that dimension one of the SIMD1 solution space can be interpreted as column sequency (feature one), complexity (feature 27), or preference (features 28 and 30) thus highlighting the problem of substitutability between features which have high intercorrelations but very different psychological meanings. Dimensions two and three cannot be satisfactorily interpreted using the present feature set.

Referring to Table 6.2 dimension one of SIMD3 can be interpreted as feature 10 (average grain) and dimension two as row sequency (feature two). Features six (squareness) and 28 (preference scale one) are strongly related to the INDSCAL solutions but they appear to be obliquely oriented to the three INDSCAL axes. Figure 6.5 shows appropriate features as vectors in the INDSCAL solution spaces for

1. Feature 34 (two-dimensional sequency) was not used in this analysis as it did not appear to have any effect in the similarity judgements aside from those effects already attributable to preference and complexity.

Table 6.1 Results for the SIMD1 INDSCAL solution regressed on each of the features.

Features	Dimension 1	Dimension 2	Dimension 3	R^2
1	-.994	.033	-.101	.921*
2	-.999	.007	.051	.395
3	-.464	-.085	.882	.126
4	-.981	-.154	-.114	.611
5	-.988	.151	.033	.773*
6	.987	-.158	.034	.875*
7	-.812	.524	-.256	.218
8	-.872	.375	-.315	.064
9	-.930	.314	-.190	.200
10	-.985	.092	-.145	.461
11	.108	.977	-.183	.044
12	.995	-.102	-.016	.329
13	-.968	-.020	.251	.563
14	.104	-.664	.741	.045
15	.579	-.466	.669	.672*
16	.701	.236	-.673	.422
17	-.877	-.192	-.440	.305
18	.008	-.330	.944	.182
19	.389	.073	-.919	.126
20	.788	.522	.325	.102
21	-.964	.257	.071	.545
22	.666	-.648	-.369	.338
23	.981	-.122	.154	.305
24	-.978	.204	-.047	.696*
25	-.988	-.008	-.151	.743*
26	.803	.154	-.575	.333
27	-.995	-.007	-.104	.946*
28	-.990	.101	-.103	.957*
29	-.787	.609	-.102	.237
30	.999	-.040	.018	.929*
31	.388	-.861	.330	.114
32	.195	.353	.915	.129
33	-.989	-.098	-.111	.354
34	.851	-.076	.520	.108

* These multiple correlations are significantly greater than zero.

Table 6.2 Results for the SIMD3 INDSCAL solution regressed on each of the features.

Features	Dimension 1	Dimension 2	Dimension 3	R^2
1	.621	.782	.056	.673*
2	-.078	.985	-.156	.851*
3	.610	-.789	-.071	.127
4	.860	.415	.296	.669*
5	-.313	.890	-.352	.554
6	-.673	-.610	.418	.970*
7	-.318	.443	-.838	.666*
8	.108	.233	-.966	.274
9	.745	-.169	-.646	.468
10	.948	-.318	.002	.776*
11	.373	-.473	-.798	.517
12	.149	-.918	.367	.756*
13	.423	-.244	-.872	.252
14	-.318	.946	.069	.504
15	-.825	.548	.137	.468
16	.178	.054	.982	.172
17	.738	.513	.438	.114
18	-.663	.154	-.732	.315
19	-.518	.448	-.729	.202
20	-.419	-.069	.906	.337
21	.500	-.265	-.824	.111
22	-.491	.799	-.348	.189
23	-.850	.523	.061	.452
24	.846	-.149	-.512	.512
25	.636	.751	.177	.797*
26	.553	-.778	.298	.430
27	.648	.761	.007	.748*
28	.642	.752	-.147	.914*
29	.920	-.392	-.020	.269
30	-.796	-.594	.120	.788*
31	.736	.644	.209	.215
32	-.022	.586	.810	.056
33	-.069	.969	.237	.444
34	-.807	-.275	.522	.729*

SIMD3.

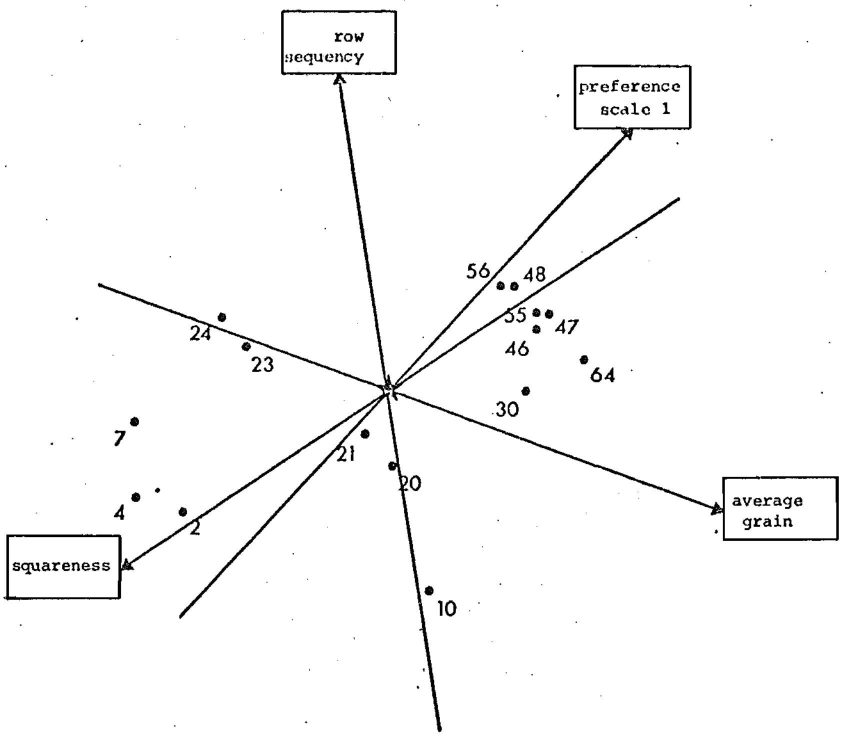


Figure 6.5 feature vectors located in dimensions one and two of the SimD3 INDSCAL solution

There is no corresponding figure for the SIMD1 INDSCAL solution as there was only one interpretable dimension in that solution.

Figure 6.5 shows four of the features in the plane of the first two dimensions of the INDSCAL solution for SIMD3. It can be seen that average grain and row sequency correspond to dimensions one and two, respectively, while preference scale one and component one are obliquely oriented to the two axes. Table 6.3 gives an interpretation of the SIMD1 and SIMD3 INDSCAL solutions in terms of the eight features used previously (in the corresponding interpretation of SIMS1, SIMS2, and SIMS3). Only one dimension of the SIMD1 solution can be interpreted, and the best explanation is provided by preference scale one. The SIMD3 solution has two interpretable dimensions, with the first dimension being a combination of squareness, average grain, and preference scale one, while dimension two can be approximated by row sequency.

Experiment SIMD8.

The procedure and stimuli used in SIMD8 were identical to those used in SIMS1 and SIMD1 except that there was a delay (ISI) of 8 seconds between the presentation of the first and second stimuli. The same analyses were run on the data as those used previously for SIMS1 and SIMD1.

TABLE 6.3 : An interpretation of the SIMD1 and SIMD3 3-D INDSCAL solutions in terms of eight of the features.

	SIMD1	SIMD3
Column Sequency	Dimension 1 $R^2 = .921$	Oblique (1-2) $R^2 = .673$
Row Sequency	-	Dimension 2 $R^2 = .851$
Squareness	Dimension 1 $R^2 = .875$	Oblique (1-2) $R^2 = .970$
Average grain	-	Dimension 1 $R^2 = .776$
Component One	Dimension 1 $R^2 = .696$	-
Preference Scale One	Dimension 1 $R^2 = .957$	Oblique (1-2) $R^2 = .914$
Complexity	Dimension 1 $R^2 = .946$	Oblique (1-2) $R^2 = .748$
2-D Sequency	-	Dimension 1 $R^2 = .729$

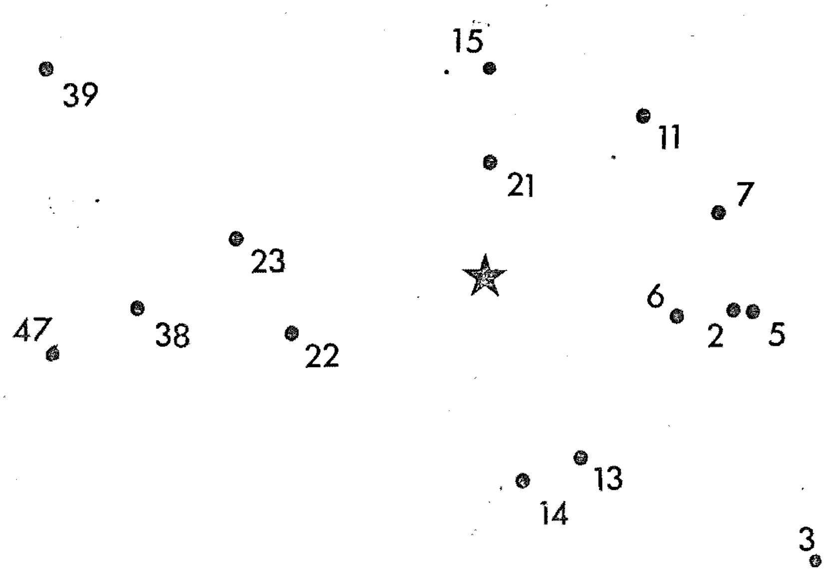


Figure 6.6 The two-dimensional INDSCAL solution for the SIMD8 results.

TABLE 6.4: Results for the SIMD8 INDSCAL solution regressed on each of the features.

Features	Dimension 1	Dimension 2	R^2
1	-.998	-.058	.903*
2	-.979	.206	.394
3	-.497	-.868	.050
4	-.986	-.168	.562*
5	-.999	.054	.801*
6	.996	-.087	.822*
7	-.941	.338	.202
8	-.996	.085	.046
9	-.794	.608	.187
10	-.987	.163	.402
11	.342	.940	.099
12	.953	-.303	.383
13	-.979	-.206	.437
14	.642	.766	.024
15	.938	-.347	.400
16	.824	.567	.151
17	-.986	.166	.261
18	.417	-.909	.075
19	.980	.197	.003
20	.925	.380	.031
21	-.942	.335	.457
22	.622	-.783	.339
23	.999	-.043	.316
24	-.990	.143	.631*
25	-1.000	.022	.731*
26	.973	.231	.160
27	-.993	-.115	.935*
28	-.999	.043	.928*
29	-.760	.649	.210
30	.999	-.032	.904*
31	.022	-1.000	.089
32	.447	.895	.221
33	-.999	.050	.354

* These multiple correlations are significantly greater than zero.

Results

Two and three dimensional INDSCAL solutions were found for the SIMD8 results. 23% of the variance was accounted for by the 2-D solutions as against only 17% of the variance accounted for by the 2-D solution. Consequently multiple regression was used to represent each of the features in the 2-D rather than the 3-D INDSCAL solution. The direction cosines between each feature and the two dimensions in the SIMD8 INDSCAL solution space are shown in Table 6.4. It can be seen that dimension one corresponds to preference or complexity while dimension two is uninterpretable. Table 6.5 gives some summary statistics for the SIMD1, INDSCAL solution. It can be seen that the fit of the INDSCAL models for SIMD1 and SIMD8 is comparatively poor. The positive correlations between the first two dimensions of the 2-D solutions suggests that there is in fact only one underlying feature. This conclusion was also suggested by the corresponding regression analyses.

Cluster Analysis

Average linkage hierarchical cluster analysis was run on single participants selected from each of SIMD1, SIMD3 and SIMD8 using BMDP1M (Dixon, 1975). The participants selected to represent the experiments were those which best met the two demands of :

(1) relatively good fit to the INDSCAL solution.

(2) relatively large saliences (subject weights) on the two INDSCAL dimensions.

Figures 6.7 to 6.9 give the configurational representations of SIMD1, SIMD3, and SIMD8, as estimated by the cluster solutions of the chosen participants embedded in the respective 2-D INDSCAL solutions. The low fit to the INDSCAL solution of the SIMD1 and SIMD8 participants is reflected in the 'messy' positioning of the clustering solutions in the two solution spaces. This 'messiness' can be seen in the tendency for there to be long and intertwined, clusters in the space. By referring back to the SIMS1 solution given in Figure 5.7 it can be seen that there is an essential equivalence between the solutions which appears to indicate that a single underlying orientation of the stimuli has been found by INDSCAL -- despite the comments made in chapter five.

The apparent changes in clustering structure between SIMS1, SIMD1, and SIMD8 will not be interpreted here as the embedded clustering solutions represent only one of the participants from each experiment. Figure 6.10 shows the SIMS3 and SIMD3 configurational representations placed together for comparison. The equivalence (in general terms) between the INDSCAL solutions is again quite striking.

TABLE 6.5: Summary Statistics relevant to the fit of the INDSCAL model for SIMD1.

	2-D Solution		3-D Solution		
Variance accounted for:	22%		25%		
Subjects' weights and individual fit	Dimensions 1	2	1	2	3
Maximum	-.52	.34	-.51	.35	.32
Minimum	-.22	.15	-.20	.10	.09
Mean	-.38	.26	-.37	.24	.21
Standard deviation	.09	.07	.09	.07	.08

r(1,2) = .42   r(1,2) = .28

r(1,3) = .08

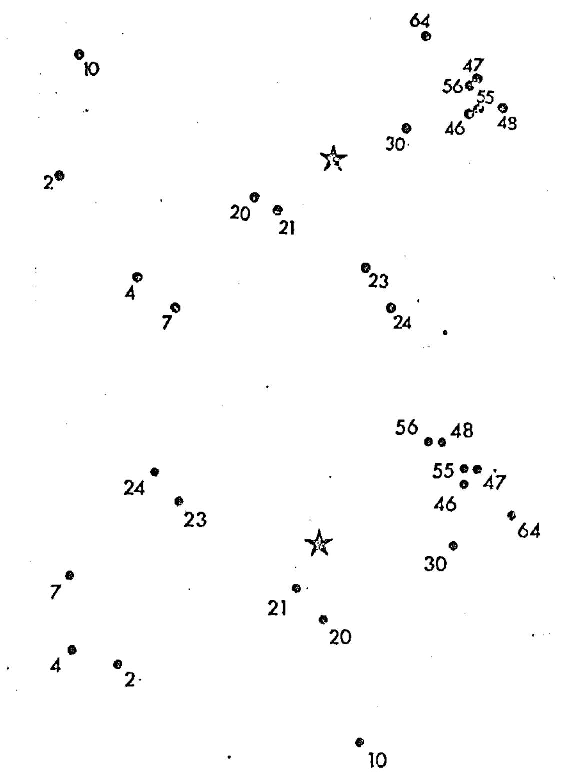


Figure 6.10 the SimS3 and SimD3 configurational representations placed together for comparison

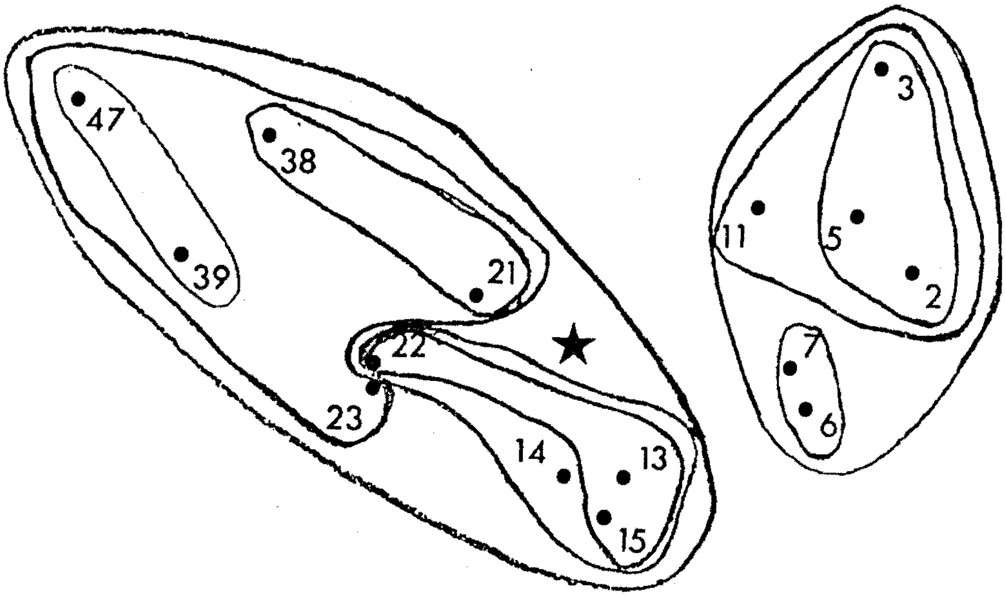


Figure 6.7 A configurational representation of the SIMD1 results.

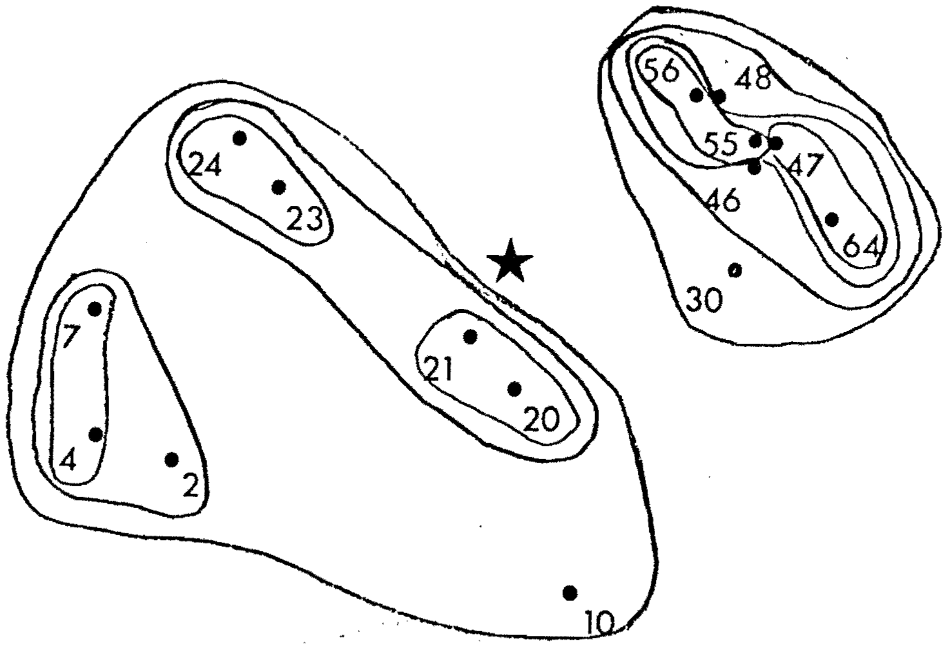


Figure 6.8 A configurational representation of the SIMD3 results.

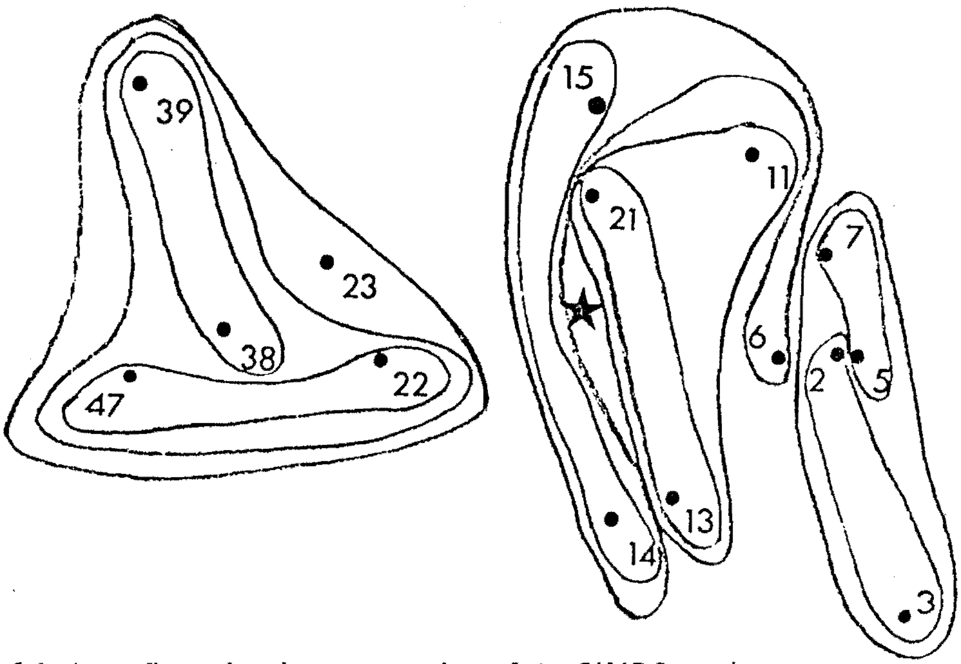


Figure 6.9 A configurational representation of the SIMD8 results.

The effect of presentation delay on geometric stimuli

The account of Posner tasks given earlier in this chapter cited studies which used verbal material (letters and words) as stimuli. The physical-/name-code distinction is not so obvious for novel geometric stimuli (such as the Walsh stimuli) where the stimuli do not have a 'name' prior to the experiment. Phillips and Baddeley (1971) used the Posner task with patterns that were too complex to allow the development of an adequate name code in the time available. The patterns comprised a 5 x 5 matrix of squares, each one of which had a 50% chance of being filled. Phillips and Baddeley (1971) found that reaction time and number of errors increased with the ISI, levelling out at 9 seconds, rather than the 1.5 seconds (for example, Posner and Keele, 1967) which is usually found with the letter matching task.

Phillips (1974) used the same procedure as Phillips and Baddeley (1971) with 4 x 4, 6 x 6, and 8 x 8 matrices. He found that retention of the most complex 8 x 8 pattern declined to the 50% chance level within 3 seconds, while 4 x 4 patterns were still well retained after nine seconds. This result suggests that the amount of forgetting of the first stimulus before the comparison in the SIMD8 task (where the ISI is eight seconds) will depend on the complexity of that stimulus. One possible strategy for making the comparison in the face of such forgetting would be to base the comparison on a single feature (such as complexity or preference) which would be retained in memory during the eight second ISI. Phillips (1974) also showed that for ISIs of less than 600 msecs, performance on the matching task was unaffected by pattern complexity only when the two patterns were in the same location. This suggests that there may be forgetting of the first stimulus in the SIMD1, SIMD2, and SIMD3 tasks as the two stimuli were projected onto two side-by-side (but not superimposed) portions of the visual field by two separate slide projectors.

The results of Phillips (1974) conform to an established pattern of reaction times for simultaneous and sequential comparisons (Egeth, 1966; Nickerson, 1967). It has been suggested that holistic, template-like processes dominate performance in sequential comparisons, whereas componential, feature-like processes dominate performance in simultaneous comparisons (Bamber, 1969; Reed, 1973). This distinction between template-like and feature-like processes appears to be somewhat similar to the distinction made in this thesis between perceptual similarity and cognitive similarity. The delayed similarity judgements used in this thesis are a first attempt to use similarity ratings (instead of or as well as reaction times) to investigate changes in the comparison process over time.

Summary

Task demands appear to affect attentional strategy by making the participant selectively attend to a single feature. The INDSCAL analyses generally gave poorer fits under the delayed conditions which, along with the results of Phillips (1974), suggest that forgetting of the first stimulus may occur during the delayed conditions. Future research on the effect of task demands on delayed similarity judgements between Walsh stimulus pairs will need to distinguish between those changes in responding which result from forgetting, and those which are due to shifts in attentional strategy (where the shifts in attentional strategy are presumably designed to minimise forgetting). The generalised similarity paradigm described in this chapter is a method of comparing and contrasting experimental procedures.

Jenkins (1979) has outlined an analogous framework for experimentation although the higher level of generality used in his treatment covers a larger range of tasks than those considered here.

CHAPTER VII

As noted in chapter one, this thesis is an attempt to combine quantitative methods with cognitive theory. The present chapter will deal with a number of issues related to the cognitive framework of this thesis. These issues are considered *en masse* in this chapter so as to minimise interruptions which would have arisen had the issues been considered in the previous chapters.

The Use of Similarity Rating Scales

Although similarity rating scales are frequently used in psychology, many researchers do not appear to consider what type of measure similarity is, when using similarity ratings. It appears to be assumed that similarities are real numbers (or integer approximations when a small number of ratings is used) on a similarity continuum. The problem is that while similarity is a particularly useful concept in cognitive psychology (as pointed out in chapters five and six), conventional treatments of similarity tend to raise conceptual difficulties in measurement theory (Gregson, 1975, 6.41, 6.42). For the present, similarity judgment will be viewed here as the mapping of pairs of stimuli onto an ordered set of fuzzy response categories. The question of whether or not there is an underlying similarity continuum from which similarity ratings may be derived by some fuzzy partition ing process need not concern us here, but the characterisation of similarities as ordered categories:

1. Puts numerically and verbally labelled similarity scales on an equivalent footing.

2. Is compatible with the notion of response strength accumulators (logogens) which underlie overt similarity responses.

Oden (1979) has shown the usefulness of a fuzzy logical model in explaining the the process of letter identification. It is envisaged here that the appropriately adapted fuzzy logical models may also prove useful in explaining the the derivation of similarity responses.

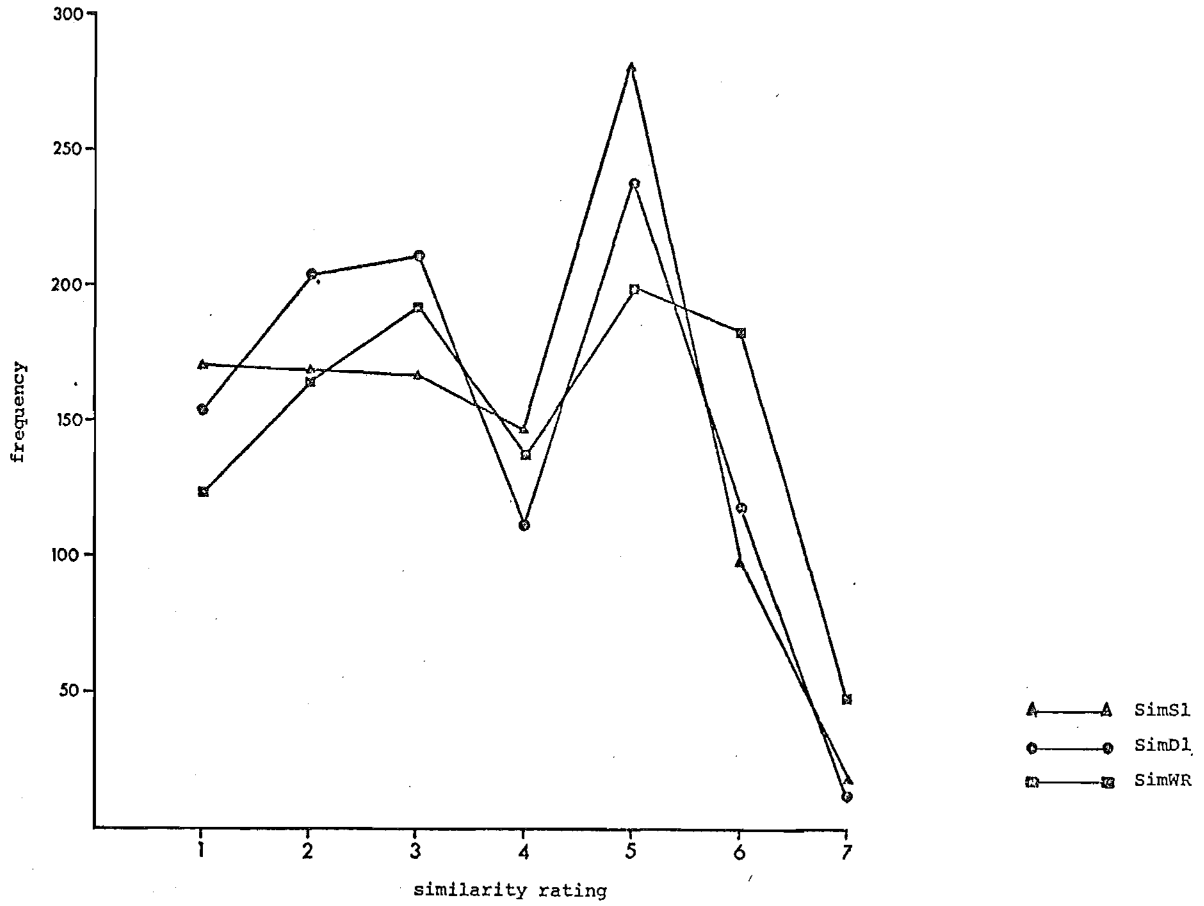


Figure 7.1 The frequency polygons of the similarity ratings (pooled across participants) for the SIMS1, SIMWR and SIMD1 experiments. It can be seen from Figure 7.1 that the SIMS1 polygon is unimodal while the SIMWR and SIMD1 polygons are bimodal.

Fagot (1979) has recently proposed a general model of relative judgment which can be applied to a variety of judgmental tasks, including similarity judgments. Fagot's model may prove to be a useful way of characterising relative judgments in psychophysical terms but it is not yet clear how compatible the model will be with cognitive notions such as feature comparison and categorising.

Reation Times in SIMWR, SIMS1, and SIMD1

The response latencies of similarity judgments do not appear to have received much attention:

"The time taken to elicit a similarity assessment has not been explored directly in recent years as for most complex reaction times it is difficult to define the precise onset and termination of the process involved. As the similarity models increase in complexity the facility with which predicitions may be made of the associated processing time correspondingly diminishes."

— Gregson (1975, p. 213)

The placement of similarity judgments within a cognitive framework (cf. chapters five and six) raises the possibility that the chronometric techniques which have recently proved to be useful in cognitive psychology (Posner, 1978; Posner and Rogers, 1978) may also be able to identify isolable systems (Posner and Rogers, 1978), psychological pathways (Posner and Snyder, 1975), and controlled processing in similarity judgments. While the application of chronometric techniques of similarity judgments is likely to be much more difficult than with same different judgments, the possibility of identifying components of the similarity judgment process certainly justifies a search for a consistent realtionship between similarity judgments and reaction times.

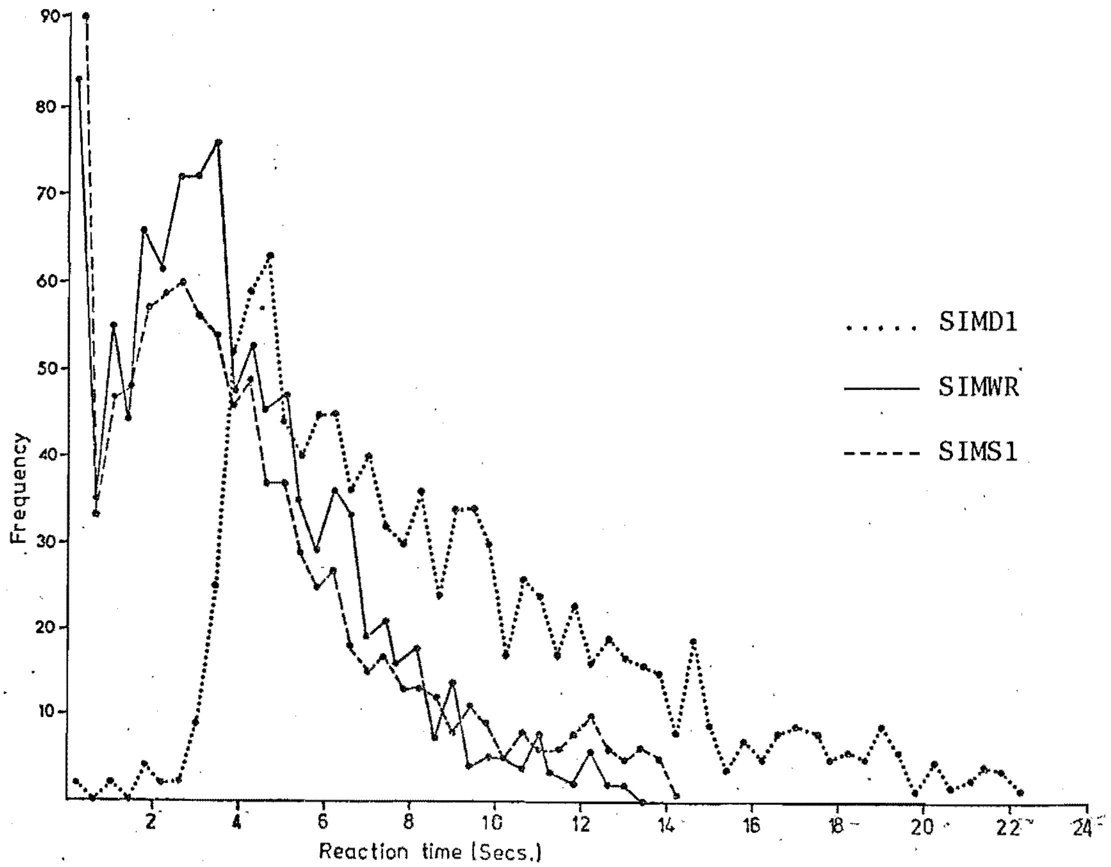


Figure 7.2 The frequency distributions of reaction times for the three experiments.

It can be seen from Figure 7.2 that the SIMD1 reaction time curve differs both in location and shape from the SIMWR and the SIMS1 reaction time curves. The early peak occuring during the first 20 msecs for the SIMS1 and the SIMWR data is due to the experimental conditions. For these experiments, the two slides in each trial were shown simultaneously for two seconds. Responses were not accepted by the computer until just before the end of the two second exposure time. Consequently, the 125 responses which occured in the first 20 msecs actually represent the sum of the responses which were made before the end of the stimulus exposure time. The presentation conditions were somewhat different for the SIMD1 experiment in that responses were accepted as soon as the second slide was presented.

The apparent difference in location between the SIMS1 and the SIMD1 reaction time distributions of 120 msecs may thus be due to the fact that the method of reaction time measurement used in SIMS1 and SIMWR has the effect of giving a participant something in the order of a 120 msec handicap in comparison with the SIMD1 reaction time method.

The SIMD1 curve has a smaller slope after the peak at 240 msecs than the other two curves after their respective peaks, implying that some of the similarity judgments take longer in the SIMD1 experiment. The peak in the SIMWR curve is larger than that of the SIMS1 curve, probably due to a smaller number of judgments being made before the end of the exposure time. Apart from this, the SIMS1 and the SIMWR curves are similar, with the possibility that the SIMWR curve lags about 40 msecs behind the SIMS1 curve.

To summarise, inspection of the distribution of reaction times for the SIMS1, SIMWR, and SIMD1 experiments indicates that the SIMWR and SIMS1 conditions elicit a similar pattern of reaction times, while the SIMD1 reaction times tend to be longer than the SIMS1 and SIMWR reaction times.

Figure 7.3 is a plot of the median reaction time for each of the 105 trials of SIMS1 against the corresponding 105 median reaction times of SIMWR. The heavy diagonal line indicates the points where the SIMS1 and SIMWR reaction times are the same, while the dashed lines parallel to it enclose the SIMS1 and SIMWR trials whose median reaction times do not differ by more than 400 msecs.

Figure 7.3 shows that any changes in reaction times from SIMS1 to SIMWR are not due to a simple (40 msec) lag induced by additional processing. The dashed horizontal and vertical lines in figure 7.3 are two superimposed axes where the origin corresponds to a reaction time of two seconds for both the SIMS1 and SIMWR experiments. Viewed in this manner, quadrant one is much less densely populated (only six of the 105 points are enclosed in it) than the other three quandrants. For present purposes the 20 msec band around the heavy diagonal line will be taken as a region where the SIMS1 and SIMWR reaction times are essentially the same.

The distribution of points in figure 7.3 suggests a tendency for increased reaction times in the SIMWR condition when the corresponding SIMS1 reaction times are shorter than two seconds, and decreased SIMWR reaction times when the SIMS1 reaction times are greater than two seconds. However, the graph also shows a correlation between medians and variances of the reaction times, which argues against a graphical comparison of reaction times elicited in the two experimental tasks.

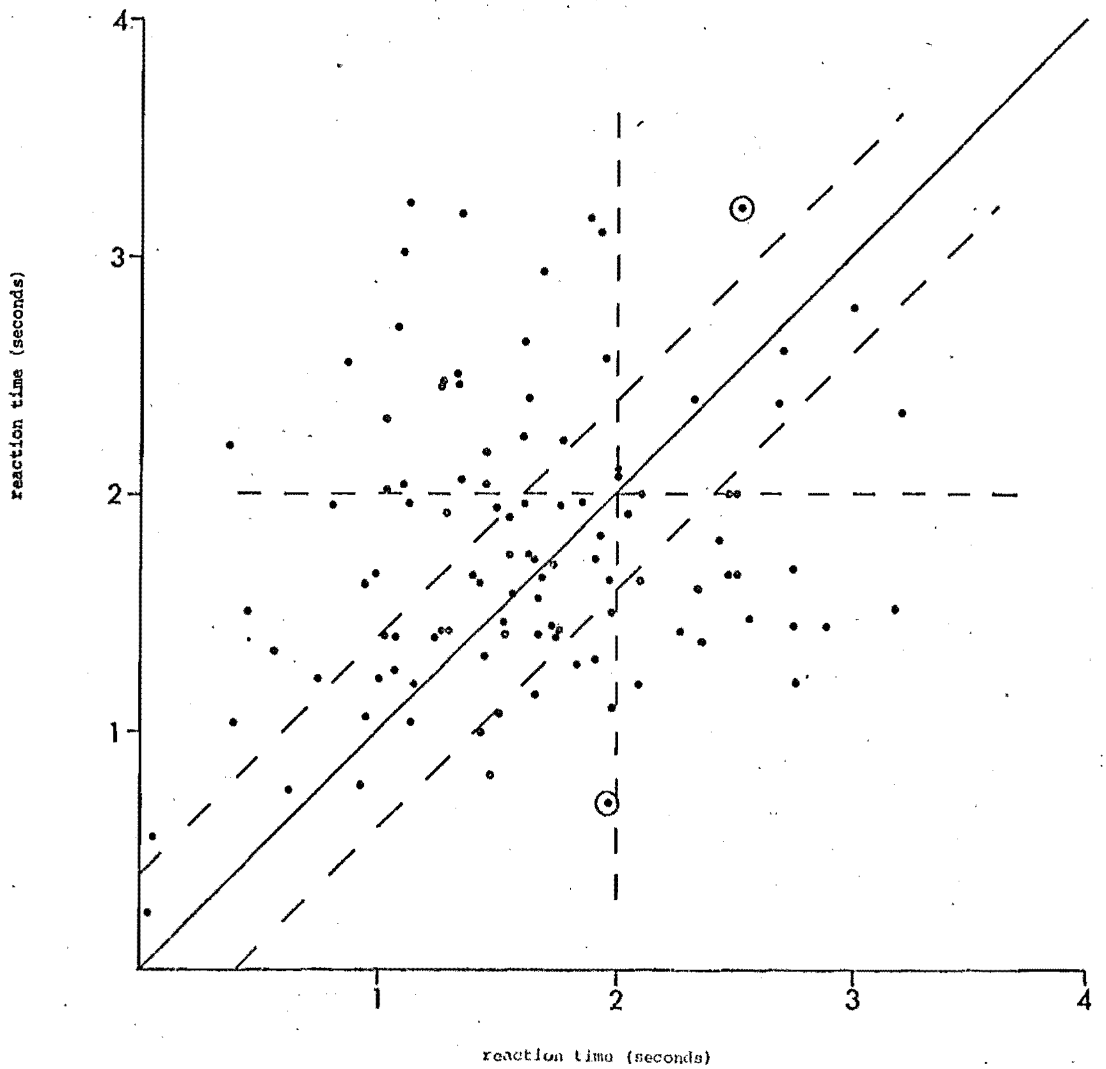


Figure 7.3 Median reaction time for each of the 105 SIMS1 trials (x-axis) plotted against the corresponding SIMWR (y-axis) median reaction times.

Trial Number	Stimuli First	Second	Reaction Time
5	7	14	2420
6	39	21	2740
10	6	23	2360
11	3	23	2260
13	15	21	2560
21	14	38	2480
26	7	2	2480
33	47	14	2500
35*	47	22	2520
37	3	21	2340
40*	6	22	3000
52*	39	7	2040
54	39	14	2100
61	5	2	3180
62	7	21	3200
65	5	23	2740
70	5	7	2080
71*	6	14	2100
72	3	14	2880
75	23	7	2500
80*	5	14	2680
83	2	38	2760
91*	6	11	2320
95*	23	14	2700

Table 7.1 The stimulus pairs which elicited median reaction times of greater than two seconds in the SIMS1 condition (they represent 24 out of the 105 trials). The asterisks indicate those pairs where the reaction time for the judgment did not appear to decrease in the SIMWR condition (in comparison with the SIMS1 reaction time).

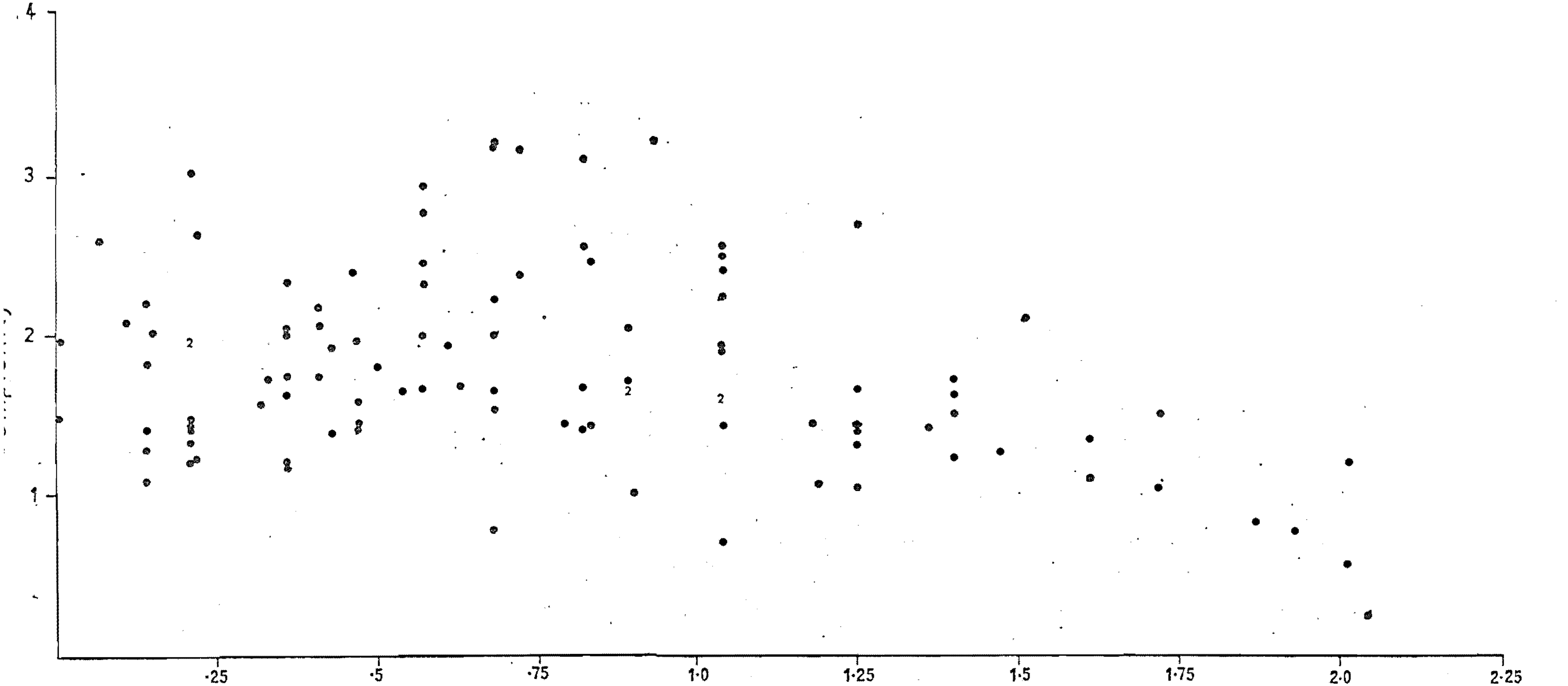
Trial Number	Stimuli First	Second	Reaction Time
2	47	21	2400
3*	38	21	2560
4	5	39	1040
7	13	22	2320
14	47	6	1660
15*	15	13	3020
17	39	3	1500
18	13	7	2020
19	23	22	2200
20	13	14	2060
22*	15	38	2460
23*	47	15	2560
25	6	13	1960
27	5	13	2180
28	15	23	1960
31*	6	15	2640
32*	5	15	2460
35*	47	22	3200
36	39	13	2040
38*	7	22	3160
39	39	47	1620
41	2	22	1340
45	23	38	1940
46*	13	2	2500
50	6	2	2240
55*	11	2	2940
57*	47	13	2700
58*	5	22	3220
63	13	23	1800
73	47	3	560
76*	47	23	3100
82*	11	21	3180
86	22	21	2040
89	2	14	1220
93	15	11	2220

Table 7.2 The stimulus composition of trials where the reaction time for SIMWR was more than 400 msecs larger than the equivalent SIMS1 reaction time. The asterisks indicate those trials where the reaction time was greater than 2.4 seconds.

It would be necessary to replicate these reaction time findings[1] before attempting to interpret them in terms of the interaction of stimulus features and attentional strategy. If the findings are replicable then a comparison of those stimulus pairs which produced relatively long reaction times in the SIMS1 condition, but not in SIMWR, with the stimulus pairs which had relatively long reaction times in SIMWR, but not in SIMS1 (table 7.2) may give detailed insight into the effect of attentional strategy on processing during a similarity judgment.

Figure 7.4 shows the relationship between reaction time and the difference in the complexity of the stimuli for the SIMWR results. There appears to be only a weak relationship in Figure 7.4. The slight curvilinearity present may actually be due to the

1. Such replication would need to be not only in terms of the overall pattern but would also need to involve similar patterns of reaction times for each stimulus pair across the SIMS1 and SIMWR conditions.



'Figure 7.4 Reaction time (x-axis) plotted against the difference in complexity for mediational effect of similarity. That is, similarity tends to have an inverted U relationship with reaction time, and similarity has a linear relationship with complexity difference.

The evidence outlined in this section suggests that analysis of reaction times may be a useful way of distinguishing the perceptual and cognitive processing required by different experimental tasks. The next section will look at the joint similarity reaction time relationship.

Changes in the Joint Similarity Reaction Time Distribution.

The SIMWR, SIMS1, and SIMD1 reaction time distributions were outlined in the previous section where it was pointed out that differences in the method of reaction time measurement make the comparison of decision times across experiments difficult, if not impossible. The reaction time distributions were segmented into seven (approximately) equally dense regions and the 7x7 similarity reaction time frequency table was constructed, for each experiment. Tables 7.3 to 7.5 are the resulting frequency tables. The informational uncertainty in bits is between 5.3 and 5.4 for all three tables. (The maximum uncertainity for a 7x7 table is 5.615 bits.) It is interesting to note that although the three tables represent a total of 3,150 responses, a similarity rating of seven (completely similar) was given only 81 times (this represents 2.6% of the responses) in the course of the three experiments.

The similarity reaction time relationship for the three experiments is represented visually in figures 7.5 and 7.6.

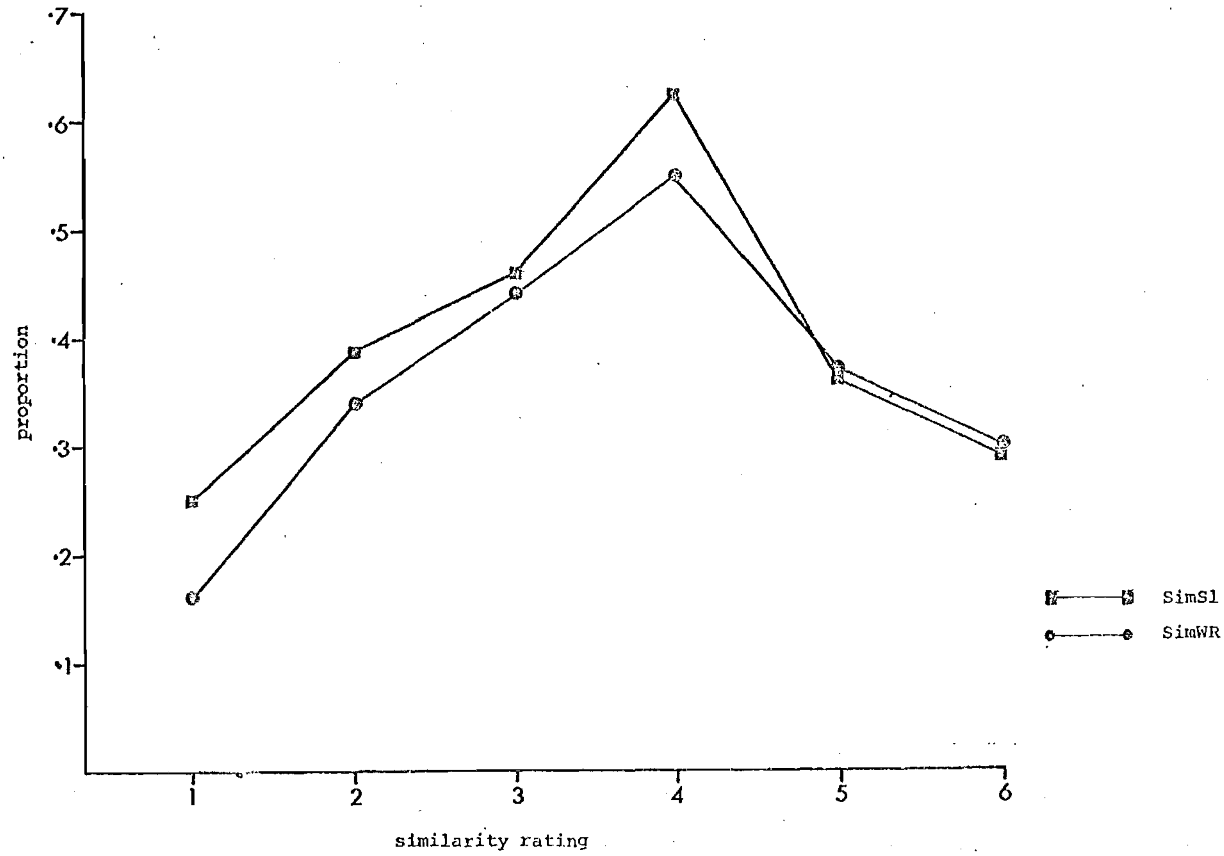


Figure 7.5. A plot of the proportion of responses made to each of the first six rating points which had latencies of more than 2.2 seconds for SIMS1 and SIMWR. This figure shows that the similarity reaction time relationships for SIMS1 and SIMWR are approximately equivalent, but that there is a tendency for a higher proportion of the ratings of one to four to be made after 2.2 seconds of measured time in the SIMS1 condition.

Table 7.3: A cross tabulation of frequencies of similarity rating x discretised reaction time for the SIMS1 experiment.

Similarity Rating	Reaction time category (msecs) 0-400	400 -1000	1000 -1600	1600 -2200	2200 -3000	3000 -4800	>4800	Total
1	48	25	27	27	11	11	20	169
2	29	31	30	14	21	24	21	170
3	12	24	27	28	20	26	30	167
4	4	13	18	17	24	37	34	147
5	47	44	51	38	42	27	32	281
6	19	15	19	17	8	5	15	98
7	4	1	3	6	2	1	1	18
Total	163	153	175	147	128	131	153	1050

Table 7.4: A cross tabulation of frequencies of similarity rating x discretised reaction time for the SIMWR experiment.

Similarity Rating	Reaction time category (msecs) 0-600	600 -1200	1200 -1600	1600 -2200	2200 -2800	2800 -3800	>3800	Total
1	47	24	16	16	7	5	8	123
2	35	26	26	22	22	17	17	165
3	23	23	26	37	27	34	23	193
4	10	13	16	22	29	26	22	138
5	22	32	22	51	20	29	24	200
6	31	44	32	22	15	23	16	183
7	11	9	6	6	7	2	7	48
Total	179	171	144	176	127	136	117	1050

Table 7.5: A cross tabulation of frequencies of similarity rating x discretised reaction time for the SIMD1 experiment.

Similarity Rating	Reaction time category (secs) 0-2.2	2.2 -2.8	2.8 -3.6	3.6 -4.6	4.6 -5.8	5.8 -7.8	>7.8	Total
1	30	31	22	23	22	16	10	154
2	39	29	27	21	29	29	30	204
3	29	26	39	39	28	28	22	211
4	3	1	12	18	17	31	29	111
5	33	31	42	42	31	31	27	237
6	23	24	19	15	16	10	11	118
7	5	6	1	1	0	1	1	15
Total	162	148	162	159	143	146	130	1050

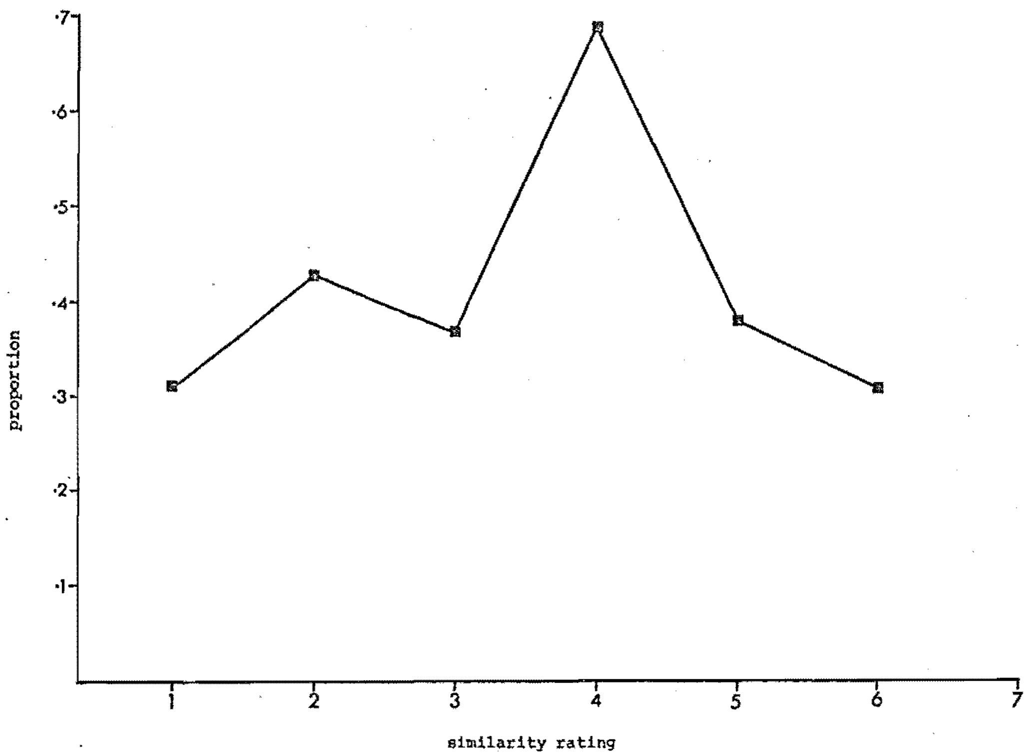


Figure 7.6 A plot of the proportion of responses made to each of the first six rating points which had latencies of more than 4.6 seconds in SIMD1.

The criterion for SIMD1 is 4.6 seconds as it includes roughly the same proportion of of SIMD1 responses as the 2.2 seconds criterion does for the SIMWR and SIMS1 responses. The higher proportion of longer latencies for the first two points on the SIMD1 similarity scale (as against the SIMS1 and SIMWR scales) may indicate that participants are biased towards judging stimulus pairs to be dissimilar when part of the stimulus information is degraded by forgetting.

Experiments SIMS1 and SIMD1 were designed so that 12 individuals participated in both experiments, with six participating in SIMD1 first. Figures 7.7 (a) and (b) show the similarity reaction time relationships for eight of these participants. The two participants on the left of each figure did SIMD1 first. The points plotted are the mean reaction times for each similarity rating, with the continuous lines (and squares) representing SIMS1 and the dashed lines (and circles) representing SIMD1. It can be seen from figures 7.7 (a) and (b) that the mean reaction times exceed six seconds for only two participants, one of whom did SIMS1 first while the other did SIMD1 first. For the other participants it can be seen that the corresponding SIMD1 mean reaction times tend to be uniformly longer than the corresponding SIMS1 mean reaction times. In the case of four of the six participants (who did not have mean reaction times of greater than six seconds) the curves are roughly parallel. Figures 7.7 (a) and (b) thus show differences between individuals where the SIMD1 mean reaction times are similar in pattern to those of SIMS1 but tend to be between one and three seconds longer, depending on the individual.

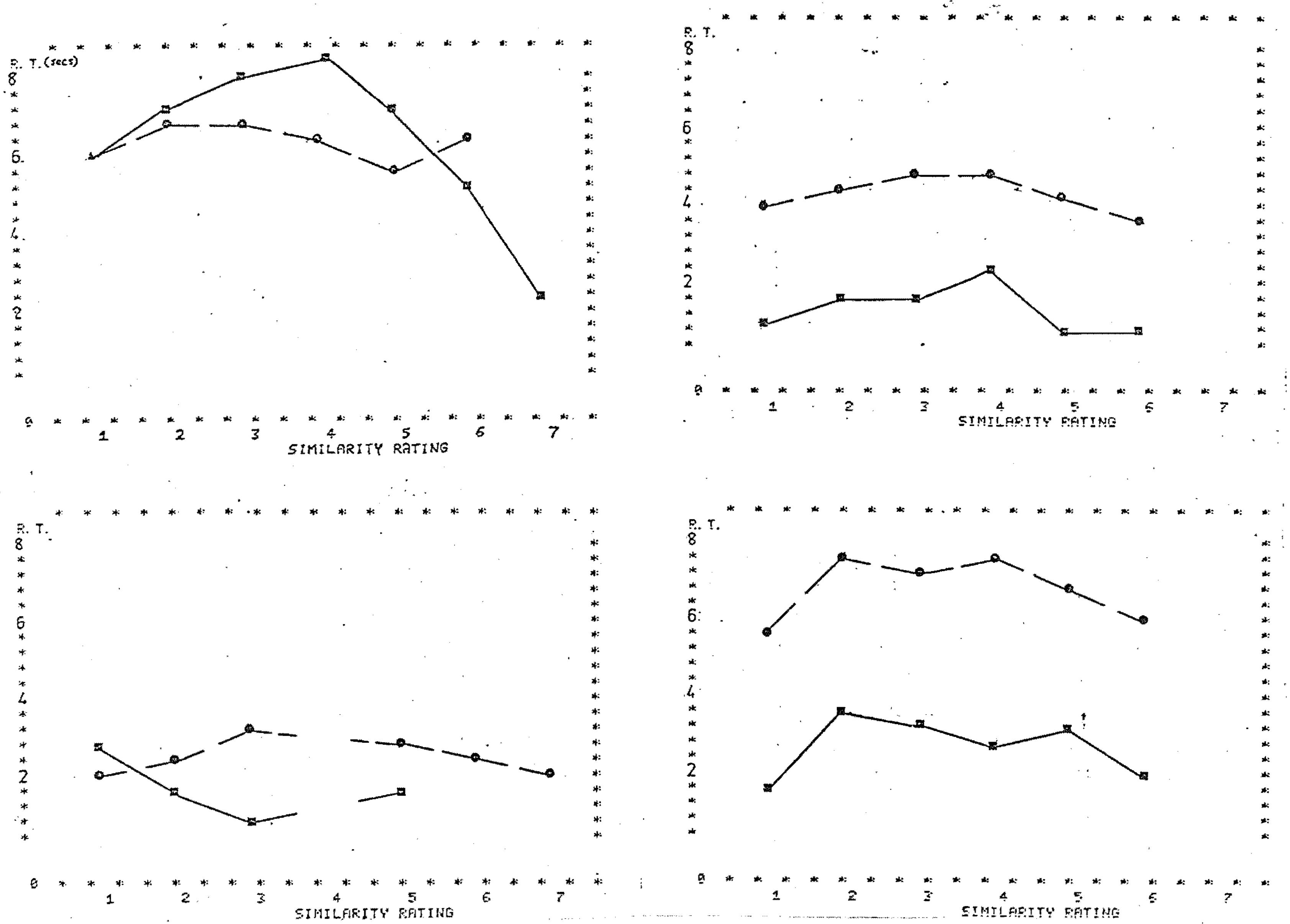
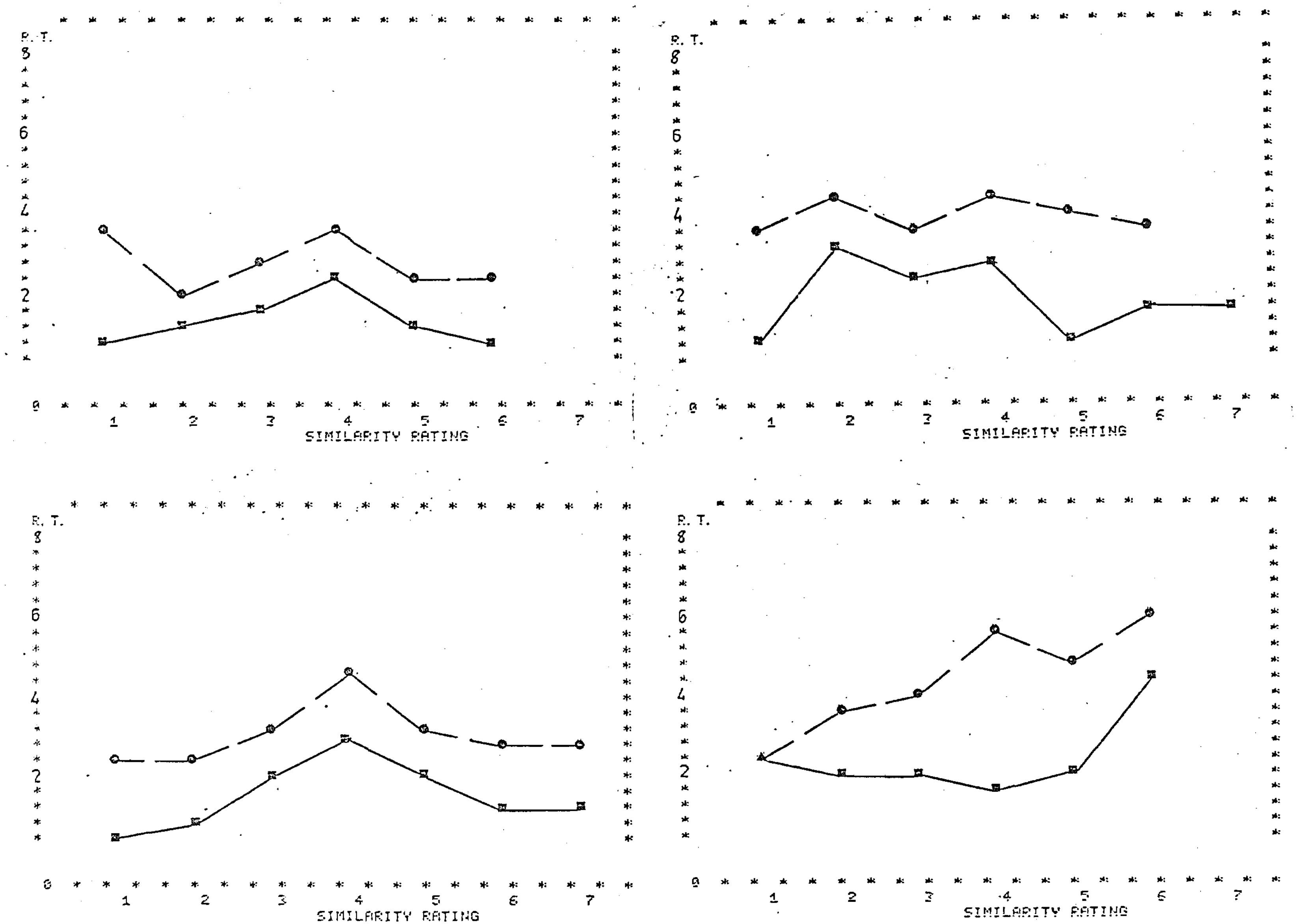


Figure 7.7a Graphs of similarity rating (x-axis) versus reaction time in seconds (y-axis) for four of the participants.

SIMS1

SIMD1



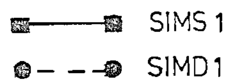


Figure 7-7b Graphs of similarity rating (x-axis) versus reaction time in seconds (y-axis) for four of the participants.

The appropriateness of the INDSCAL model

The INDSCAL model was used extensively in chapters four, five and six where there was evidence both for and against its appropriateness. For instance, the analyses in chapter five found some, but not all, of the INDSCAL dimensions to be interpretable in terms of physical features.

The results of the INDSCAL analyses in chapter six showed the INDSCAL solutions to be stable across different amounts of delay in the presentation of the second stimulus. This stability suggests that the INDSCAL method is in fact recovering a unique representation of stimulus subsets one and three. A comparison of the clustering analysis and INDSCAL results shows both agreement and disagreement between the two methods in the similarities implied between the stimuli.

One method for checking the validity of the INDSCAL model is to examine the co-ordinates of points in the subject space (Kruskal and Wish, 1978, p.61). If there are substantial negative values this may mean, for instance, that the dimensionality is too large for the data. Such an examination of the co-ordinates of points in the subject spaces (for the 2-D INDSCAL solutions) of SIMWR, SIMS1, SIMS2, SIMS3, SIMD1, SIMD3, and SIMD8 showed that only SIMD1 and SIMD8 had negative values. (The co-ordinates were consistently negative on the first dimension of both the SIMD1 and SIMD8 solutions, as shown in tables 7.6 and 7.7). This accords with the suggestion made in the previous chapter that the SIMD1 and SIMD8 judgements were generally based on a single feature.

	Dimension 1	Dimension 2
1	-.524	.161
2	-.028	.307
3	-.409	.199
4	-.337	.340
5	-.367	.296
6	-.393	.296
7	-.374	.346
8	-.437	.245
9	-.339	.301
10	-.222	.256
11	-.349	.153
12	-.507	.243

Table 7.6 Dimensional saliences from the INDSCAL solution to the SIMD1 results.

	Dimension 1	Dimension 2
1	-.474	.287
2	-.439	.229
3	-.367	.301
4	-.178	.404
5	-.222	.292
6	-.451	.264
7	-.496	.283

Table 7.7 Dimensional saliences from the INDSCAL solution to the SIMD8 results.

In summary, then, it appears that the INDSCAL model is generally appropriate as far as the direction of the group space is concerned, but that the INDSCAL model may not be relevant for some participants despite its apparently good fit.

Further clustering analysis.

Figure 5.7 showed the clustering solution for a SIMS1 participant (the one with the highest correlation between observed and predicted scores) embedded within the 2-D SIMS1 INDSCAL solution. The results of two other SIMS1 participants were also cluster analysed (again using BMDPIM). The first participant had a correlation of .79 and dimensional weights of .60 and .47. Her clustering results are presented as a dendogram in figure 7.8 as it was impossible to embed the clustering results in the INDSCAL 2-D solution. (Note in particular how stimulus 5 clusters with 13 and 21 while stimulus 6 clusters with 14, 15, 22, and 23.

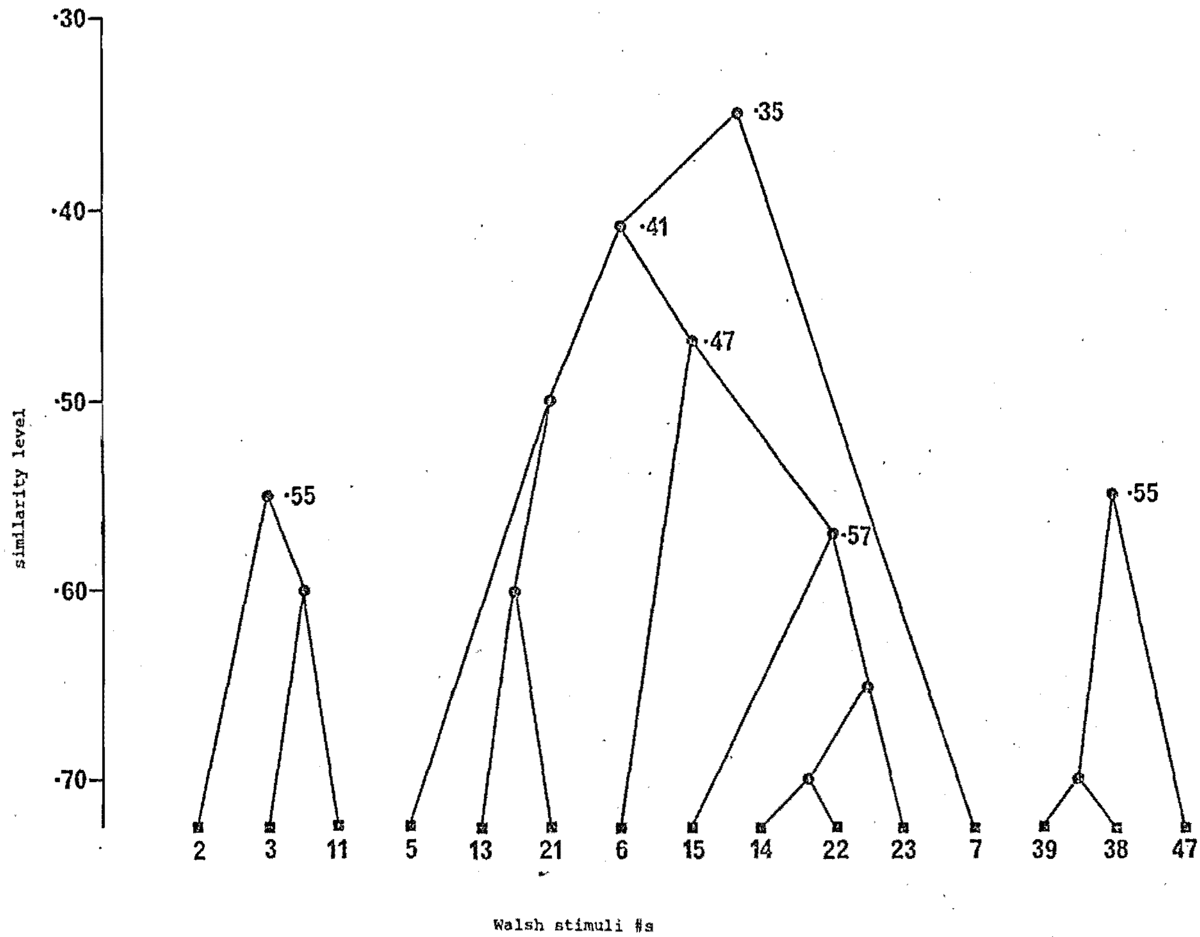


Figure 7.8 The clustering solution obtained from the results of one of the SIMS 1 participants, presented as a dendogram.

The second participant whose results were used in this subsidiary analysis had a correlation of .84 and dimensional weights of .73 and .35. In contrast to the other two participants whose SIMS1 results were cluster analysed (remembering that one participant's analysis has already been described in Chapter five), the cluster analysis results were able to be embedded in the INDSCAL solution without any major distortion to the implied pairwise similarities as shown in figure 7.9.

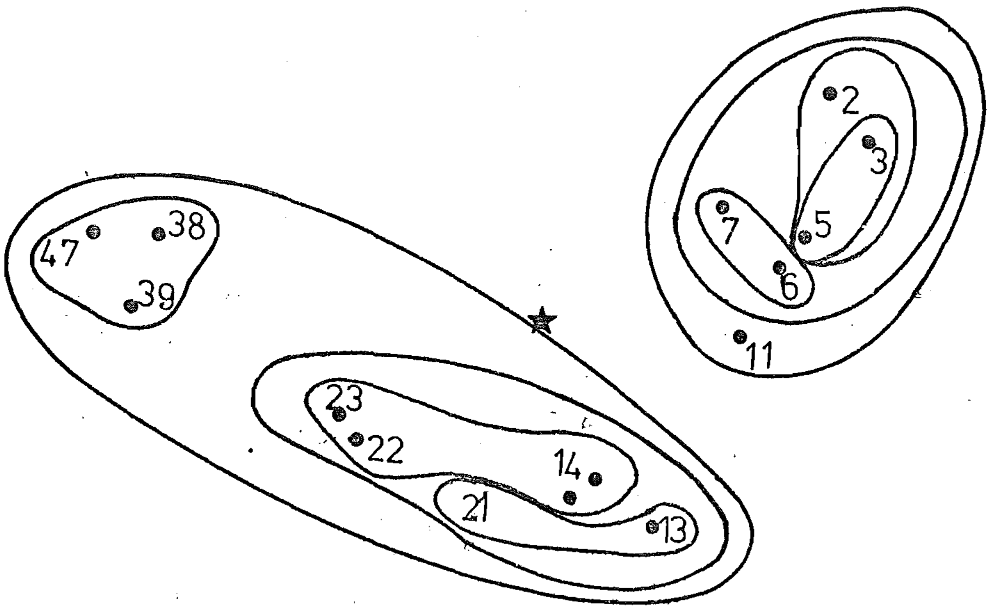


Figure 7.9 The configurational representation obtained from the results of one of the SIMS1 participants.

Figure 7.9 shows that stimulus No. 11 is generally seen by this person as dissimilar to the rest of the Walsh stimuli. (The average similarity rating between the cluster entities was only 3.2 when 11 joined the 2, 3, 5, 6, 7 cluster.) However, an examination of the raw data indicated that the perceived similarity between stimulus No. 11 and stimuli 3, 13, and 15 respectively was five (on the seven-point rating scale). This indicates that the clustering analysis may be inducing its own kind of distortion on the data due to violations of the ultrametric axiom (Johnson, 1967).

Average-linkage heirarchical cluster analysis (the method of clustering used in this thesis) is only one among many clustering methods. Clustering methods in general have a number of well known limitations (Cormack, 1971). However, cluster analysis may be legitimately used when the purpose of the analysis is exploratory (as in this thesis) rather than confirmatory (Chignell and Stacey, 1980). INDSCAL is an analytic method that has proved its worth in a number of empirical studies (Wish and Carroll, 1974). The present results indicate that it may be possible to cross-validate the INDSCAL model for individual participants by attempting to embed the results of an appropriate form of cluster analysis in the INDSCAL solutions.

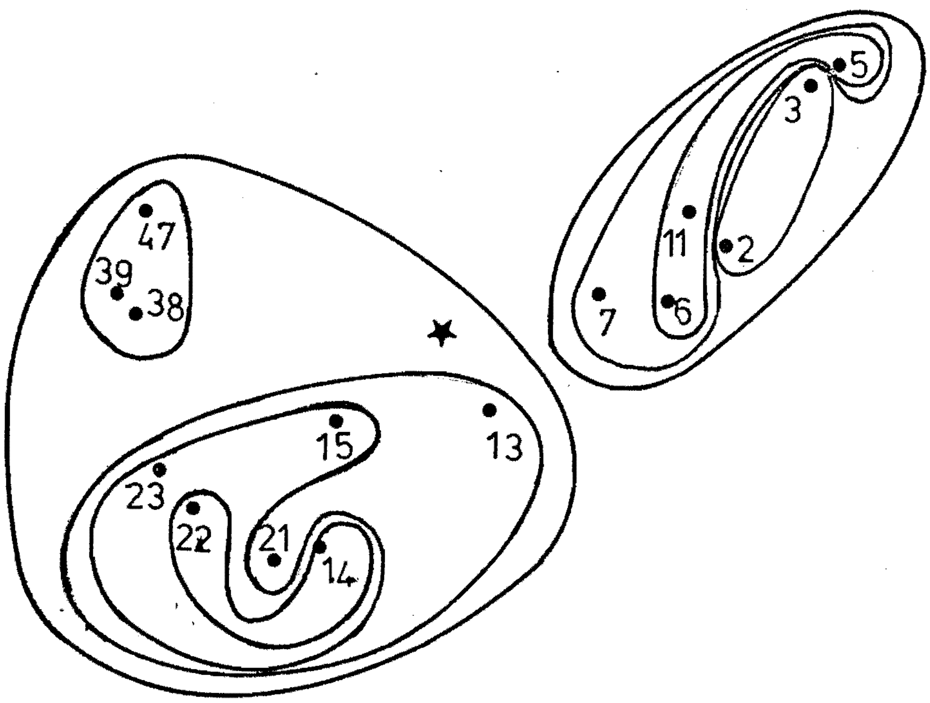


Figure 7.10 A configurational representation of the SIMWR results using the clustering results of one of the SIMWR participants.

A comparison of figures 7.9 and 7.10shows that these two versions of the configurational representations for SIMS1 and SIMWR[1] are roughly equivalent. These results, taken with those in chapter six, suggest that the INDSCAL group space may in fact yield a unique representation which is not changed by the selective attention instructions and presentation delays used here. The clustering analyses, on the other hand, appear to be very sensitive to individual differences and no attempt will be made here to estimate changes in cluster structure across experiments. A comparison of the various cluster analyses of the SIMWR, SIMS1, SIMD1 and SIMD8 data does, however, indicate that the stimuli in set one appear to be generally perceived as being composed of three clusters. Stimuli 38, 39, and 47 are in one cluster, stimuli 2 and 3 are in another, while stimuli 13, 14, and 22 are in the third. The clusters to which the other seven stimuli in the set belong appear to vary across individuals and tasks. The basic clustering for the set one stimuli appears to be that which is shown in both figure 7.9 (the clustering results for a SIMS1 participant) and figure 7.10(the clustering results for a SIMWR participant). This clustering is as follows:

1. In theory, at least, there should be a configurational representation for each participant which will be the appropriate cluster analysis embedded in the transferred INDSCAL group space.

Cluster one: 2, 3, 5, 6, 7, 11

Cluster two: 13, 14, 15, 21, 22, 23

Cluster three: 38, 39, 47

The preservation of the order of the stimulus numbers between clusters shows that Walsh ordering may be having a strong effect on similarity judgements.[1]

Some relevant issues in Quantitative Modelling

One of the problems associated with the comparison of configurational representations across experiments is the variability in clustering results across individuals. Shepard and Arabie (1979) have developed the ADCLUS method which can be viewed as a discrete analogue of principal components analysis. In effect it locates binary features in the data and assigns weights to these features which can be taken as estimates of the relative salience of those features. Carroll and Arabie (1979) have produced a three-way generalization of the ADCLUS model which they have called INDCLUS. In the INDCLUS model each participant has a separate weight for each subset (cluster) although the number and composition of the subsets will be the same for all participants (cf. the INDSCAL group and subject spaces). The INDCLUS algorithm was not available to the author at the time of writing but it appears to be a method that might produce clustering solutions which are stable across different experiments. If such stability occurs, then it may be possible to use the subject weights for the respective INDCLUS solutions to measure changes due to experimental tasks, although this approach did not work for the INDSCAL subjective weightings derived in the experiments reported in chapters four, five and six.

Another analytic method which may prove useful is fuzzy clustering which is outlined in Appendix D.

At present the field of quantitative modelling (and especially the area which is relevant for the analysis of psychological data) is in a state of flux. The situation is exacerbated by there being a number of models which appear to be compatible with cognitive metatheory, but which lack a formulated statistical structure or even an approximate set of statistical

1. The number of a Walsh stimulus is derived from its position in the sequency-ordered matrix of Walsh stimuli and is obtained by counting up the columns of this matrix. The first column (column sequency = 1) is numbered one to eight the second column is numbered nine to sixteen, and so on.

guidelines obtained from Monte Carlo simulation.

The methods which are likely to prove useful within a framework such as that outlined in the present thesis are:

INDSCAL (Carroll and Chang, 1970)
INDCLUS (Carroll and Arabie, 1979)
Three-mode Scaling (Tucker, 1972)
Covariance Structure analysis (Joveskog, 1978).

The information gained from these methods could also be supplemented and more closely investigated using regression analyses.

CHAPTER VIII

The information (knowledge) contained in stored representations, and the manner in which this information is stored, is an important and unresolved issue in cognitive psychology. One of the foci of this issue has been the representational nature of mental imagery. Anderson (1978) has argued that it is not possible to use psychological experimentation to dist inguish between pictorial and propositional storage of information.

Kosslyn's (1978) computer graphics metaphor is, however, one way of accounting for the phenomenon of visual imagery which is compatible with the theory outlined in chapters one and four. According to the computer graphics metaphor:

> ". . . a visual image . . . is thought of as a 'surface representation' generated from some underlying 'deep representation', The relationship between a visual image and its underlying deep representation is posited to be like the relationship between a display on a cathode-ray tube and the computer program (and data) that produces it."
>
> —Kosslyn (1978, p.179)

In the present view, images are schemata activated within the system of stored representations in response to (image producing) task demands. The question remains, however, as to what form the stored representations (Kosslyn's 'deep representations') take.

Rumelhart and Ortony (1977) have given a plausible account of stored representations of knowledge. One of the advantages of their 'Active Structural Network' formulation is that it explicitly relates the functional content of information (Palmer, 1978, p.300) to the stored representation of that information.

Nelson (1977) has outlined the nature of functional information content as follows:

> "The function or relationship of the object or event that identified it as novel, interesting, or important can be specified as the initial *meaning* of the concept for the child. Elsewhere (Nelson, 1974), I have referred to this as the functional core of the concept. The core includes relations of the thing or event in time and space and to self, others, and other objects. . . Here, the term functional does not imply conventional usuage. . . Rather, it implies function from the point of view of the individual forming the concept. . . Thus, it may include what the object does, the set of its possible actions, as well as what can be done to and with it and the results of actions on it."
>
> —Nelson (1977, p.230)

She then goes on to say that the process of conceptualisation (knowledge acquisition) consists of establishing conceptual frameworks (these are analagous to the scripts of Schank and Abelson, 1977), generating functional cores (defined above), and finding identifying attributes (perceptual features). Nelson based her discussion on the cognitive development of preschool children but the following quotation also applies to adults who are dealing with objects that cannot be identified by name:

"The spatial and temporal frame or script enables the child to predict which things and events to expect and in what order. So long as these expections are met she can act according to a pre-established automatic script; she does not have to give special attention to the order of events or use problem-solving skills to figure out what the situation implies. The functional core concept enables the child to predict which things to expect in a given context, and also what to expect of a given thing when it is encountered. The identifying attributes enable the child to predict the functions of a thing before they are observed or experienced. . .(N)ew instances may be recognized wherever they are encountered by virtue of their identifying attributes."

—Nelson (1977, p.230)

While Nelson's (1977) theory is adopted here as the best available account of concept (knowledge) acquisition, only a small portion of it is relevant to the acquisition of knowledge about the Walsh stimuli. This is because the Walsh stimuli appear to lack affordance structure[1], or functional core (as explained in chapter three of this thesis). It will be argued in this chapter that the Walsh stimuli are seen as a subset of the set of possible checkerboard stimuli[2] by virtue of their underlying row and column sequencies (cf. the final sentence in the preceding quotation). An intuitively appealing illustration of this type of process is given by Norman (1969):

"If shown a new word — this chapter emphasizes the mantiness of memory — we know that the word is new. This ability to scan rapidly through memory and reject a novel item has some strong implications about the method of reaching the stored addresses of material which is in memory. . . Thus (it may be that)[3] we can say that the address of a word can be determined entirely by its physical characteristics. Note the peculiar implication; even things which we have never experienced before (such as some of the Walsh stimuli)already have a specific memory location reserved for them. This is not to be taken as a statement concerning

1. The notion of affordance structure is outlined by Gibson (1977).

2. There are definable, mathematically and empirically, checkerboards which are not Walsh stimuli (e.g., Smets, 1973. Chipman, 1977) and participants will have experienced some elements of this set.

3. The two bracketed portions have been added to the original quotation by myself.

prior knowledge.; it is only a description of the sensory orientation of memory."

—Norman (1969, p.162).

The remainder of this chapter will use sorting tasks in attempting to demonstrate the perception and cognition of Walsh stimuli as particular instances of the stored concept of black and white checkerboard patterns.

Sorting the complete set of Walsh stimuli

Four sorting experiments were carried out to investigate the underlying logical structure of the Walsh stimuli as perceived by the participants. The experimental procedure for each of the sorting tasks will be described separately, and then the results of the four experiments will be presented and discussed. The sorting experiments each used a different set of ten participants, making a total of 40 participants, none of whom were used in any of the experiments reported previously in this thesis.

Sort 1

The first experiment (Sort 1) was a free sort (that is, there were no restrictions on the number of piles or the number of stimuli placed within a pile) using the 64 Walsh stimuli plus an additional stimulus which was a completely white square (analagous to the homogeneously black Walsh stimulus in the bottom left-hand corner of figure A.1.).

Sort 2

The second sorting task was a free sort on the 64 Walsh stimuli (the white stimulus was not added).

Sort 3

The third sorting task also used the 64 Walsh stimuli, but in this experiment the participants were required to use seven (and only seven) piles.

Sort 4

The completely black stimulus (stimulus No. 1) was removed from the set of 64 Walsh stimuli. Participants were required to carry out a free sort on the resulting set of 63 stimuli.

Results

The four conditions (Sort 1, Sort 2, Sort 3, and Sort 4) were run to check how much small changes in stimulus context and the nature of the sorting task (unrestricted versus seven-pile sorting) would effect the sorting results. The method of analysis to be described below is similar to that used by Miller (1969), among others. The sorting results were converted into four proximity matrices (one for each sorting experiment) by summing the number of participants who placed the two stimuli in the same pile, for all possible stimulus pairings. These proximity measures were then divided by 10 (the number of participants in each experiment) to give a decimal fraction. The four proximity matrices were then analysed using average linkage hierarchical clustering implemented with the BMDP1M programme (Dixon, 1975).

Figure 8.1 gives a tree representation of the Sort 1 clustering solution. The tree for this experiment, and those for the other three sorting tasks, show only those clusters which have an average proximity of .35 or greater between their members. Table 8.1 gives the clusters which are apparent in figure 8.1.

Cluster Number	Cluster Composition					
1	1	2	3	4	5	6
	7	8	16	24	32	40
	48	56				
2	9	10	17	18		
3	11	12	13	14	15	25
	33	41	49	57		
4	19	20	21	22	23	26
	34	42	50	58		
5a	27	28	29	30	35	36
	37	38	43	44	45	46
	51	52	53	54		
5b	31	39	47	55	59	60
	61	62	63			

Table 8.1 A clustering interpretation of the Sort 1 results.

Figure 8.2 gives a tree representation of the Sort 2 solution. In general, there is less evidence of clustering than that shown in figure 8.1. This was in spite of the fact that the median number of piles used in Sort 2 was 7.5, against a median of 11.5 piles in Sort 1.[1] However, the clustering that is apparent in the Sort 2 results is generally compatible with the Sort 1 clustering solution, as shown in table 8.2.

1. The smaller the number of piles used, the more likely it is that two stimuli will be placed together thus the expected proximity of a randomly selected pair of stimuli should be higher.

Cluster Number	Cluster Composition	Equivalent Sort 1 Cluster
1a	1 2 3 4 5 6 7 8	1
1b	16 24 32 40 48 56	1
2	9 10 17 18	2
3	11 12 13 14 15	3
4	19 20 21 22 23	4
5	25 33 41 49 57	3
6	26 34 42 50 58	4
7a	27 31 59 63	5a & 5b
7b	28 35 36	5a
7c	29 30 60 61 61	5a & 5b
7d	37 38 39 43 44 45 46 47 51 52 53 54 55	5a & 5b

Table 8.2 A clustering interpretation of the Sort 2 results.

Figure 8.3 gives a tree representation of the Sort 3 solution.

Table 8.3 shows that the Sort 3 results are similar in structure to the Sort 1 results.

Cluster Number	Cluster Composition	Equivalent Sort 1 Cluster
1a	1 2 3 4 5 6 7	1
1b	8 16 24 32 40 48 56	1
2	9 10 17 18	2
3a	11 12 25 33 41 49 50 57 58 59	3 & 4
3b	13 14 15 22 23	3 & 4
4	19 20 21 26 34 42	4
5a	28 29 35 36 37 43 44 45	5a
5b	30 31 38 46 47 51 52 53 54 55 60 61 62	5a & 5b
5c	27 63	5a & 5b

Table 8.3 A clustering interpretation of the Sort3 results.

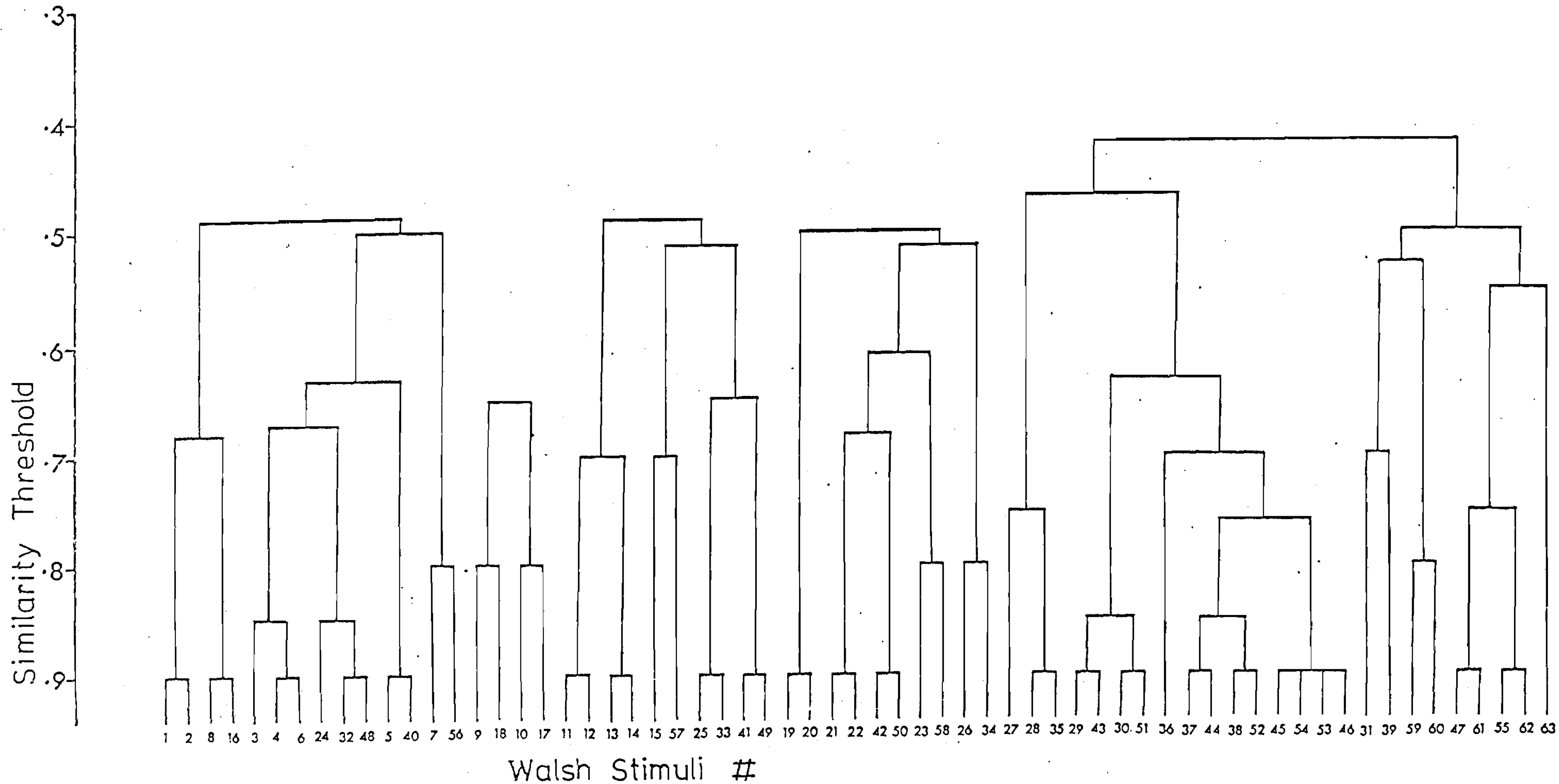


Figure 8.1 A tree representation of the Sort1 clustering solution.

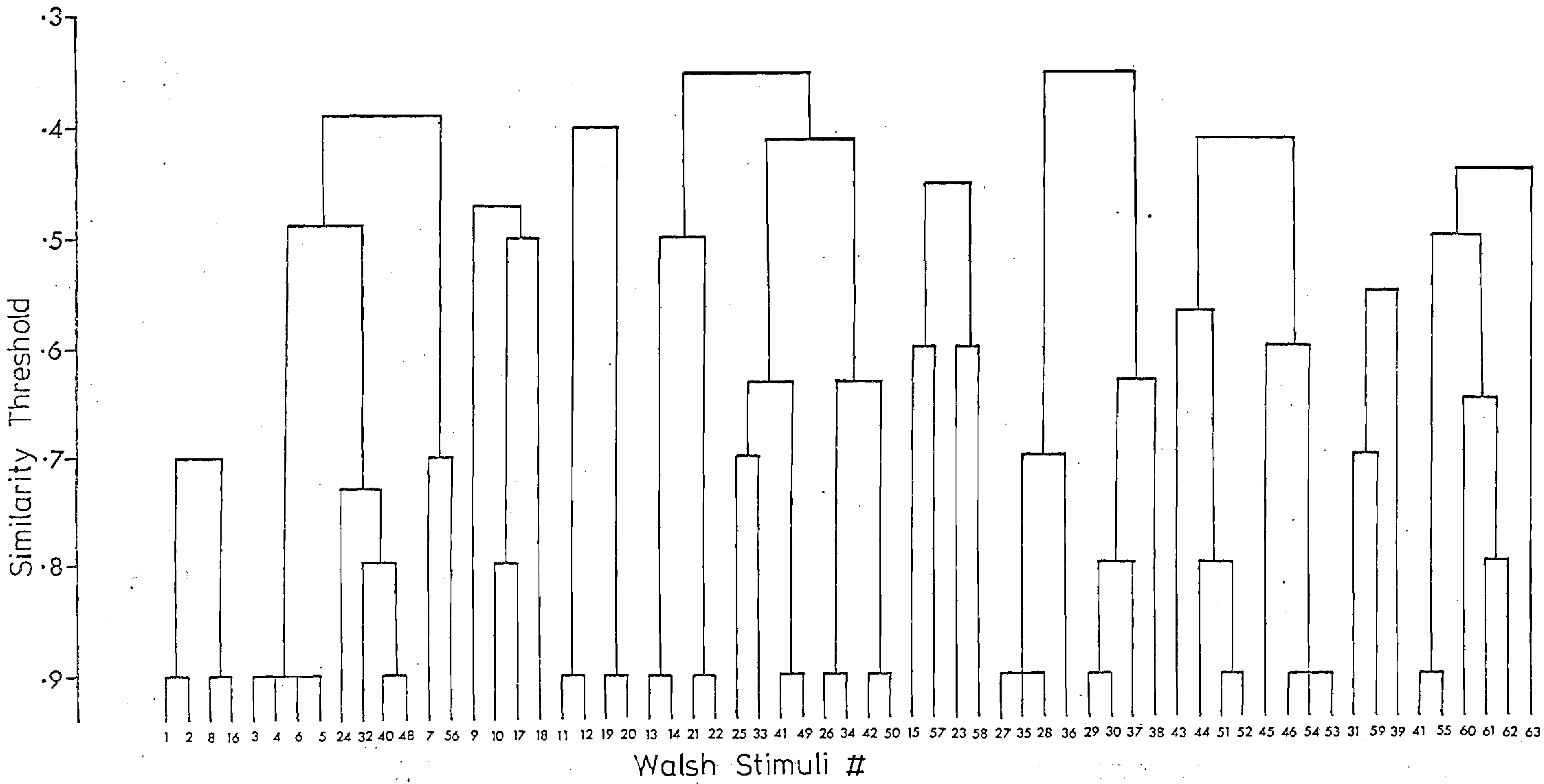


Figure 8.2 A tree representation of the Sort2 clustering solution.

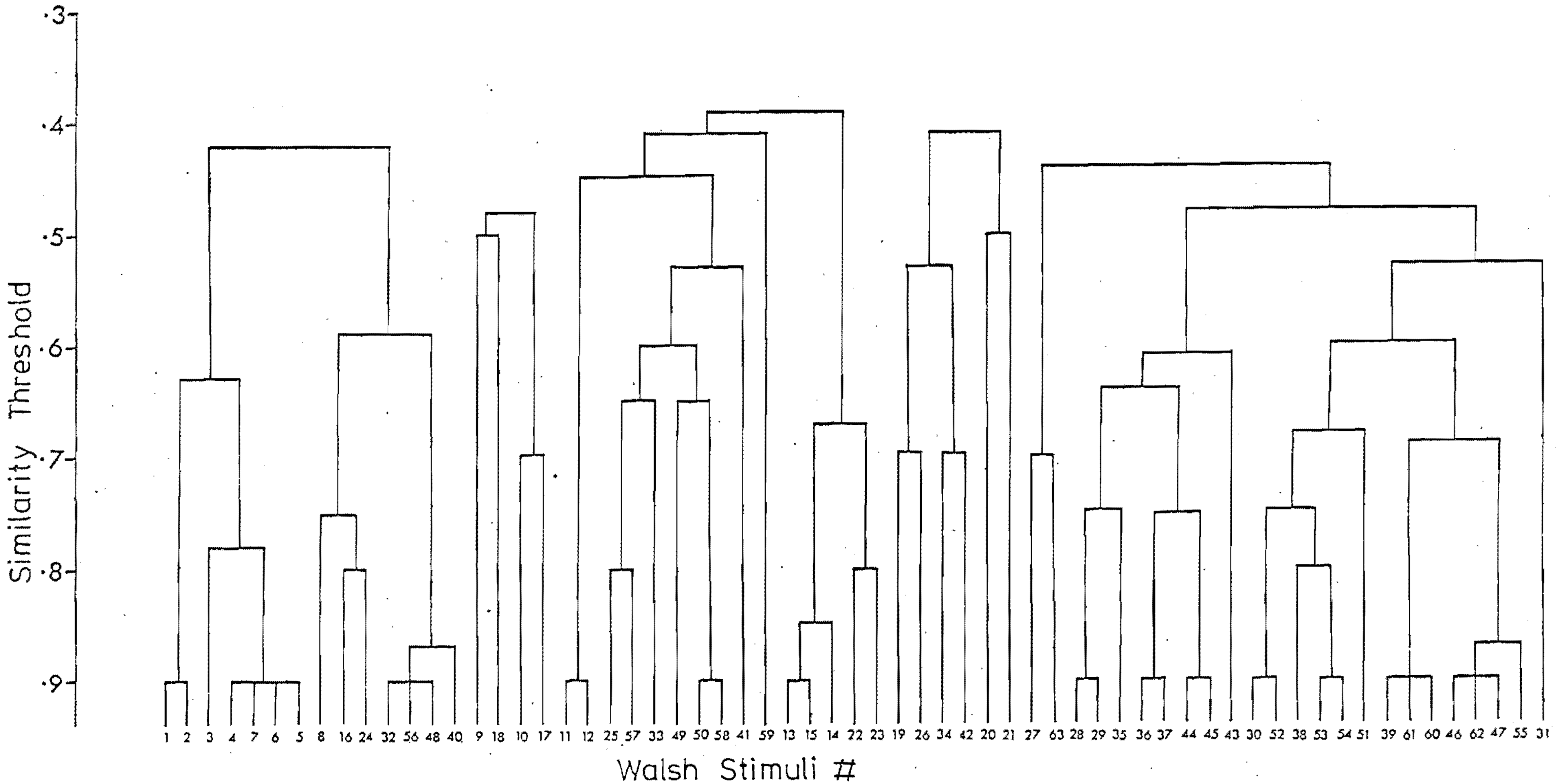


Figure 8.3 A tree representation of the Sort3 clustering solution.

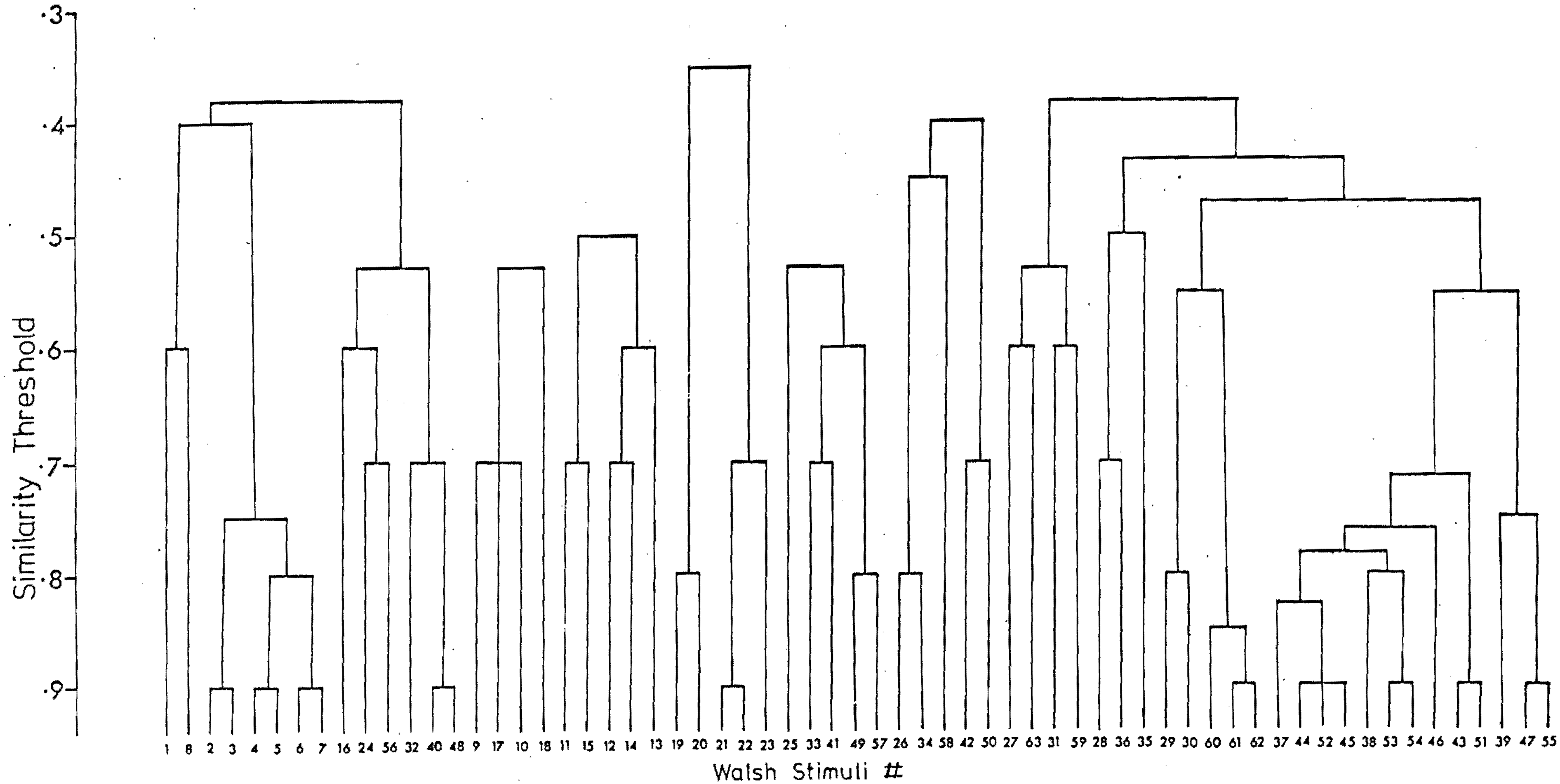


Figure 8.4 A tree representation of the Sort 4 results.

Figure 8.4 gives a tree representation for the Sort 4 clustering solution. Table 8.4 gives the corresponding interpretation in terms of the Sort 1 results. It can be seen that the clustering structure in the Sort 4 results is weak, and only roughly comparable with the Sort 1 results.

Cluster Number	Cluster Composition						Equivalent Sort 1 Cluster
1	1	2	8	16			1
2a	3	4	5	6			1
2b	24	32	40	48			1
2c	7	56					1
3	9	10	17	18			2
4	11	12	19	20			3 & 4
5a	13	14	21	22			3 & 4
5b	25	33	41	49			3
5c	26	34	42	50			3
6	15	23	57	58			3 & 4
7a	27	28	35	36			5a
7b	29	30	37	38			5a
8a	43	44	51	52			5a
8b	45	46	53	54			5a
9	31	39	59				5b
10	47	55	60	61	62	63	5b

Table 8.4 A tree representation of the Sort 4 clustering solution.

The Comparison of Sorting Results

One problem that arises in the type of experimentation reported above is that of how to identify structural commonalities and differences between different sets of sorting results:

"(One) problem encountered in the use of sorting is in comparing and summarizing the results from different experiments. For many psychological studies, this problem is dealt with through the use of statistical distributions and measure of central tendency. But from semantic studies using the method of sorting, questions about 'central tendency' cannot at present be asked, much less answered. There simply is not basic quantitative datum derived from sorting which is comparable between experiments, and hence there is no evidence that repetitions of experiments give results which converge on any particular finding."

—Boorman and Arabie (1972, p.247).

One structural measure outlined by Boorman and Arabie (1972, pp.234-237) was the pair bond height measure. A method for calculating this measure between two sorts is to first calculate the incidence matrix M(P) = [Aij] nxn for each partition P, where

$$a_{ij} = \begin{cases} 0 \text{ if } e_i, e_j \text{ are in different cells of P.} \\ 1 \text{ if } e_i, e_j \text{ are in the same cell of P.} \end{cases}$$

$D(P_1, P_2)$ — the pair bond height measure is then the number of pairwise distinct entries in the incidence matrices M (P_i), normalised by division by n(n-1). Table 8.5 gives the values of 1 — D for each of the pairs of partitions.

	Sort 1	Sort 2	Sort 3	Sort 4
Sort 1				
Sort 2	.880			
Sort 3	.866	.862		
Sort 4	.900	.916	.892	

Table 8.5 The proportion of entries in the pairs of incidence matrices which were equal (i.e., both 0 or both 1).

This measure is the proportion of entries in the two incidence matrices which agree. As can be seen there is about 90% agreement between the cluster interpretations derived from the four sorts. An alternative method for investigating structural consistencies between sorts will be considered in a later section of this chapter.

Discussion of the first four sorting tasks

The presence of the two extra stimuli in Sort 1 appeared to produce a strong clustering structure. This may have resulted from increased homogeneity in responding across participants. It was suggested earlier that the Walsh stimuli will be perceived as particular instances of the stored concept of black and white checkerboard patterns. The presence of the all-black and the all-white stimuli in the stimulus set may increase the awareness of the checkerboard nature of the Walsh stimuli through the process of contrast (Sherif, Taub, and Hovland, 1958). Of the four clustering solutions, the one for Sort 1 best reflects the checkerboard structure of the Walsh stimuli, as shown below.

The clusters obtained from the Sort 1 data (given in table 8.1) can be interpreted in terms of the numbers of vertical stripes or horizontal bands that they contain. Cluster one consists of all the Walsh stimuli which contain eight horizontal bands, plus the remaining stimuli which do not have any vertical stripes. Cluster two contains the four stimuli with either two or three vertical stripes, and no more than two horizontal bands. Cluster three contains all the remaining stimuli which satisfy at least one of the following two conditions:

1. The stimulus contains no horizontal bands,
2. The stimulus has two vertical stripes.

Cluster four consists of all those remaining stimuli which have three horizontal bands and/or three vertical stripes. Cluster 5-b contains those stimuli remaining which have seven horizontal bands and/or eight vertical stripes. The sixteen stimuli in cluster 5a are all the possible combinations of stimuli with three, four, five, or six horizontal bands, and four, five, six or seven vertical stripes.

Figure 8.5 gives a visual interpretation of the Sort 1 clustering results in terms of the sequency-ordered matrix of Walsh stimuli. Thus, it can be seen that the Sort 1 results generally indicate that the participants in this experiment organised the Walsh stimuli in terms of the number of horizontal bands, and the number of vertical stripes.

A comparison of the results for Sort 2 and Sort 3 shows the effect of restricting the number of piles which can be used in the sorting task. The Sort 3 results show more evidence of clustering that the Sort 2 results, and the Sort 3 clusters are more interpretable than the clusters derived from the Sort 2 results. Comparing the Sort 2 and Sort 3 results, it appears that requiring the participant to use a particular number of piles (as in Sort 3), where the number of piles is close to the number of categories into which the stimuli are expected to be divided, will produce more homogeneity in participants' responding.

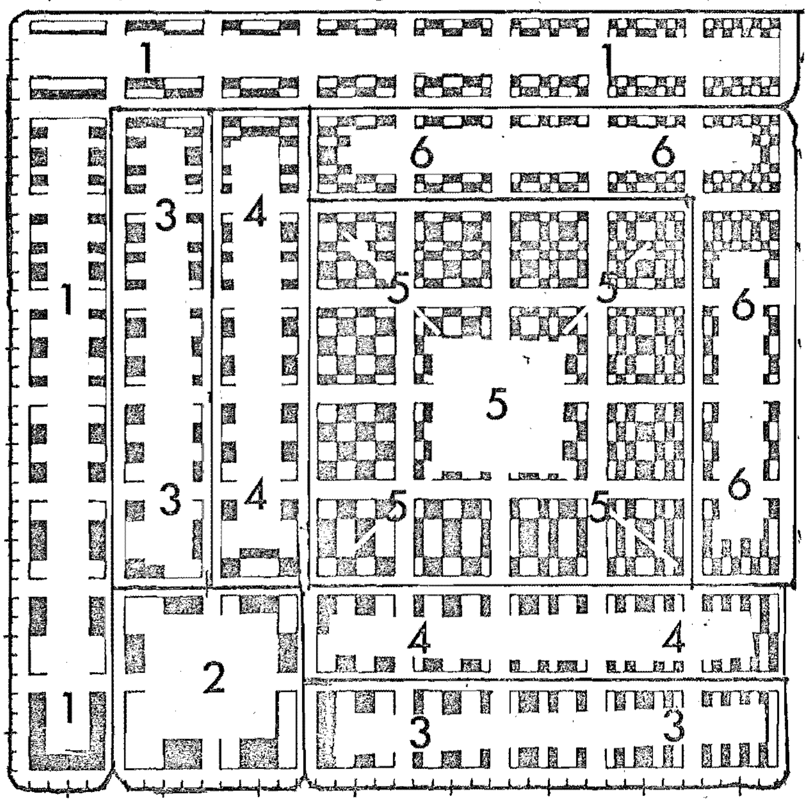


Figure 8.5 A visual interpretation of the Sort 1 clustering results in terms of the sequency - ordered matrix of Walsh stimuli.

Comparison of the proximite matrices

Comparison of the partitions derived from cluster analysis of the grouped sorting data (see above) suggested a high level of agreement between the Sort 1, Sort 2, Sort 3, and Sort 4 sortings. This kind of comparison can be criticised because it compares the derived clusters, rather than using the more direct approach of comparing structure within the proximity matrices.

Hubert (1979) has developed a general concept of concordance which can be used to compare the structure of proximity matrices. As Hubert and Levin (1976) have noted, the most obvious measure of relationship between two proximity matrices is a Pearson product-moment correlation between the corresponding N(N-1)/2 off-diagonal entries of the two matrices. This is the measure that will be used here. The method of Hubert (1979) may be used in future sorting st udies with the Walsh stimuli to check for the structure implied by the present results.

The results of the correlational analysis of the proximity matrices are given below.

Results

Table 8.6 gives two summary statistics for the proximities derived from the four sorting tasks.

	Mean	Standard deviation
Sort 1	3.18	1.70
Sort 2	1.51	1.76
Sort 3	2.46	2.14
Sort 4	2.27	1.97

Table 8.6 Means and standard deviations of the Proximities Derived from the Four Sorting tasks.

The intercorrelations between the proximity matrices are given in Table 8.7.

	Sort 1	Sort 2	Sort 3	Sort 4
Sort 1				
Sort 2	.754			
Sort 3	.767	.633		
Sort 4	.766	.610	.766	

Table 8.7 Intercorrelations between the off-diagonal elements of the four proximity matrices.

Table 8.7 indicates that there is a certain amount of common structure in the four proximity matrices derived from the sorting tasks, which supports the results obtained by finding the proportion of matches (generally about 90%) between the incidence matrices which were reported in a previous section.

The general problem of detecting common structure between different proximity matrices is recognised here, although the psychological interpretability of the present results, plus the degree of apparent concordance found in the previous analysis, go some way towards establishing the validity of the present clustering interpretation.

Summary

The results of the four sorting tasks reported above indicate that the Walsh stimuli tend to be organised in terms of their underlying sequences (the number of vertical stripes and horizontal bands that they possess). There was also a 'tightening up' of the apparent clustering structure when the all-black and the all-white stimuli were introduced. It was suggested in a previous section that this 'tightening up' was due to a contrast effect. Another way of interpreting this result would be as an adaptation level effect on a set of relative similarity judgements (where the terms relative similarity is used in the sense of Gregson, 1975). It is possible for instance, that in allocating a stimulus to a pile a relative similarity judgement is made between the similarity of the stimulus to the pile being considered and the similarity of the stimulus to the other stimuli not in the pile. The presence of extreme stimuli (such as the all-black and the all-white stimulus in the Walsh stimulus set) would make the other (non-extreme) stimuli appear to be (pairwise) more similar to each other. As an example of this type of process, London and Edinburgh may seem quite distant in the context of Glasgow and Brighton, but comparatively close in the context of New York and Beirut.

Sort 5

An additional experiment (Sort 5) was also run, where the participants made a free sort of the 35 stimuli used in the similarity experiments reported in chapters five and six of this thesis. The pattern of results was similar to that of Sort 1, with the stimuli being distinguished in terms of underlying sequency, although there were marked individual differences (in terms of their sorting results) between the 25 participants of Sort 5.

The sortings of the Walsh stimuli reported in this chapter may be extended in a number of ways. Firstly, the sorting of other (non-Walsh) checkerboard stimuli as well as mixtures of Walsh and non-Walsh checkerboard stimuli should test more explicitly the suggestion made here that the Walsh stimuli are perceived as a subset of the stored concept of black and white checkerboard patterns.

Secondly, the apparently consistent and interpretable preference, similarity, and sorting structures of the Walsh stimuli found in this thesis may be tested for their interrelationships. If the pattern of results reported here is found to be stable under replication then it may well be worthwhile attempting to relate the perception and cognition of the Walsh (and other checkerboard) stimuli to measurable intellectual and personality variables within the individual.

CHAPTER IX

Aims and Theoretical Orientations

As stated at the beginning of Chapter One, the overall aim of this thesis is to develop experimental paradigms and methods of analysis which would be suitable for the quantitative modelling of the perception and learning of visual stimuli. The development of such paradigms required a theory which would relate knowledge gained through perception to the behavioral consequences of that knowledge. In the present thesis it has been assumed that some form of stored representation is used to incorporate past experiences into present perception. The references cited in Chapter One show this to be a common assumption in current cognitive theory.

The first step in devising experimental paradigms which would explore the acquisition and utilisation of stored representations is to choose a set of stimuli which are both novel and quantifiable. It was decided not to use verbal (language-related) stimuli as these would tend to be associated with overlearned stimuli in the environment of the subjects. It was assumed that learning would result in a change in the pattern of responding which was elicited. The Walsh stimuli used in this thesis are quantifiable, but not ecologically valid (in the sense of Neisser, 1976). The tradeoff between ecological validity and quantifiability has been characterised in Appendix A.

A number of quantifiable visual stimuli have been used in previous studies of visual form (e.g. Attneave & Arnoult, 1956; Zusne, 1970). The Walsh masks, however, provide a counter instance for various models of similarity, as shown in Appendix A. In addition, the Walsh masks are representable by a terse analytic formula, and they are a closed set so that the transformation from one stimulus to another is also representable. Other sets of stimuli which have been used, or which readily suggest themselves as empirically suitable material, do not have all of these properties, although they may have some of them.

The use of novel and non-verbal stimuli meant that it would have been precipitate to specify stimulus features and then incorporate these features as factors in an analysis of variance design. Consequently, considerable effort has been devoted to specifying (1) what the physically, or mathematically, definable properties of the Walsh stimuli are (Chapter Two), and (2) what the psychologically salient features of the Walsh stimuli are (Chapters Three, Four and Five). The contradistinction between these two stages of the analysis is fundamentally important because it leads to consideration of feature extraction and selective attention as processes which have to be modelled. The length of this feature extraction process should not be surprising in view of the general difficulty in isolating the metrics of visual form (Brown & Owen, 1967; Zusne, 1970).

The possible role of preference or selective attention following some evaluative role, and its theoretical relationship with adaptive fitting to the environment was considered on pages 33 to 35 of this thesis. Perceptual learning in general may also be seen as being closely related to the process of adaptive fitting to the environment:

"Perceptual learning, then, refers to an increase in the ability to extract information from the environment as a result of experience and practice with stimulation coming from it. That the change should be in the direction of getting better information is a reasonable expectation, since man has evolved in the world and constantly interacts with it."

— Gibson, E. J. (1969, p.3).

In the situation where the environment consists of a novel set of visual stimuli presented in the psychological laboratory, a process of adaptation will still occur, but this process will be sensitive to some aspects of the experimental procedures used, particularly those which impose demands on the subject. This view is incorporated in the notion of task demand effects which was considered in Chapter Four. Introductions to task demand effects in psychophysics and cognitive psychology are given in Helson (1964) and Schneider & Shiffrin (1977), respectively.

Feature Extraction

The psychologically salient features of the Walsh stimuli were determined in two stages. (1) a set of candidate features based on quantifiable properties of the Walsh stimuli were developed (feature extraction); (2) candidate features were evaluated in terms of how well they could account for the experimental responses (feature selection). In the feature extraction stage additional candidate features were derived by using conceptual ranking to scale preference and complexity. A total of 24 quantifiable properties of the Walsh stimuli are outlined in Chapter Two (pp. 14-15 & 27-28) and in Appendix B.

The first experiment reported in this thesis (pp. 16-21) was a pilot study to investigate the usefulness of direct scaling of subjectively perceived stimulus properties using verbal rating scales. The use of loosely constrained verbal labels was designed to encourage the participants to use their own, rather than the experimenter's, criteria in judging the stimuli. The experiment, referred to as E1, yielded uninterpretable results in terms of criterion identification, but was reported in order to show the extent of individual differences in numerical responding which may arise when using a rating method.

It is possible that unambiguous definitions of each rating scale and training of participants in the general use of rating scales may have yielded better behaved judgmental scales:

"... (T) he making of magnitude estimates depends on an arbitrary association between the numbers of the judgmental scale and the physical values of the stimuli to be judged. The nature and extent of the subjects' training, with this kind of task and with the experimental stimuli, will obviously have a great deal of effect on the reliability and meaning of judgments that are made."

— Dember (1960, p.104).

The following two scaling experiments (E2 and E3) used the conceptual ranking method and in both cases interpretable results were obtained. A single scale of complexity (pp. 23-24) and six scales of preference (pp. 36-44) were identified using the conceptual ranking technique.

One problem associated with complete ranking methods in general is that they force complete transitivity on the participant's responses, whereas partial ranking, of which paired comparison is the limiting case, does not. This enforced transitivity may, conceivably, obscure the fact that a person is ranking according to more than one attribute (this could imply some intransitive pairwise comparisons). In the present case it is argued that complexity and preference (for a given person at a given time) should correspond to single attributes and that any intransitivities will result from errors in responding. In this sense, therefore, complexity and preference (relative value) are higher order constructs which may be representable as combinations of perceptual features.

A second problem arises when considering what ranks should be assigned to the stimuli following a conceptual ranking. One method (cf p.22) is to assign arbitrary numbers which obey the rule of having each row and column ranked in ascending order. Alternatively, Monte Carlo simulation may be used (as in Appendix C) to derive the expected ranks associated with each position in a conceptually ranked grid.

Once rankings have been obtained they can be used as order statistics which describe the distribution of the feature which has been ranked:

"We are reinforced in our inclination to consider tests based on ranks ... by the fact that the ranks are invariant under any monotone transformations of the variables. Any such transformation will also leave the hypothesis of independence invariant, and the ranks are therefore natural quantities to use. We still have not settled which functions of the ranks are to be used as our numbers ... the simplest obvious procedure is to use the ranks themselves ..."

— Kendall & Stuart (1973, vol. 3, p.494).

The strength of the relationship between ranks and variates depends on the nature of the underlying distribution from which the variates are sampled. The strength of this relationship can be quantified using the product moment coefficient of correlation. Exact formulae for calculating this correlation between variates and their ranks have been derived for different distributional assumptions (Kendall, 1970, p.125). In experiment E3 for instance, 35 Walsh masks were used as stimuli. The correlation between 35 variates and their ranks is .972 if the variates are uniformly distributed and .950 if the variates are normally distributed.

The most important result, however, is that the correlation between n variates and their corresponding ranks can be expressed as a ratio of the limiting value of the correlation coefficient as n tends to infinity, i.e.,

$$C_n = (n-1)/(n+1)^{.5}.C$$

where C_n is the correlation for sets of n stimuli and C is the limiting value of the correlation. This result indicates that ranks can be used without too large a loss of precision, where there are more than 30 stimuli, say, and where the underlying distribution fulfills minimal conditions such as that of having a finite variance.

Thus, most of the error in the values assigned to the ranks inferred from a conceptual grid will arise almost entirely from error in the procedure of estimating the ranks, and error in the derived scale values (i.e. the ranks themselves or summary satistics based on them) will also reflect only this ranking error. In the conceptual ranking procedure final positions in the grid are not independent of the starting configuration. Monte Carlo simulation has shown that there is a range of possible ranks symmetrically distributed around the expected (mean) rank for each position in the conceptual grid. Thus the mean ranks obtained from Monte Carlo simulation are unbiased estimators of the scale values of the stimuli placed in those positions. These mean ranks are tabulated in Appendix C for several two-way ranking configurations.

The unbiasedness of the scale value estimators means that rank estimates may be pooled over subjects who are homogeneous with respect to the scale in order to give more accurate estimates of the scale values, as was done in Chapter Three. The error free norm which this procedure closely approximates is the non-permuted ranks of the stimulus scale values.

In psychological experimentation, this error free norm corresponds to the results of an ideal (hypothetical) subject only (since people typically show marked variability in their judgments) and cannot be directly obtained by any currently available data collection methods.

In its present form, the reliability of scale values estimated from conceptual ranking depends on the number of sets of results which are pooled and the mean and variance of the pairwise inter-correlations of those sets. In this thesis, the results were pooled across relatively small numbers of subjects. Despite this, the complexity and preference scales derived from conceptual ranking were interpretable and were able to account for much of the variation in similarity responding.

The apparent robustness of the scales derived from the conceptual ranking method as currently defined justifies the search for a more rigorous formalisation of the properties of the rankings that it produces. In practice this formalisation implies regressing the ranks onto measures of stimulus properties that are independently definable.

Feature Selection

Multidimensional scaling is one method for identifying salient features. While the dimensions of a solution configuration may be regarded as defining a set of features, it is customary to justify or interpret the dimensions of a solution in terms of known properties of the stimuli, usually physical, or hedonic. Chapters Four, Five and Six attempted to interpret a number of multidimensional scaling solutions in terms of the candidate features derived in Chapters Two and Three.

Experiments SIMS1, SIMS2 and SIMS3 collected pairwise similarity comparisons for three overlapping subsets of the Walsh stimuli. Two- and three-dimensional solution spaces were fitted to the similarities data obtained in these experiments using the INDSCAL programme (Carroll and Chang, 1970).

A method outlined by Kruskal and Wish (1978) was then used to interpret the INDSCAL dimensions in terms of the set of candidate features. It was found that the INDSCAL solutions to the SIMS1, SIMS2, and SIMS3 results generally reproduced (though somtimes in rotated form) one or more from a small subset of the candidate features. The interpretation of the three solutions in terms of these features is given in table 5.7 (p.94).

The feature selection process resulted in the identification of five salient features (column sequency, row sequency, squareness, average grain, and complexity). The INDSCAL solutions could also be interpreted in terms of preference scale one and 2-D sequency, but these latter features are highly correlated with complexity.

Task Demands and Selective Attention

Figure 1.2 (p.12) in Chapter One is an attempt to provide a visual representation of the kind of theoretical orientation which was adopted in this thesis. One of the implications of that figure is that task demands directly affect attention and effort. Following a brief discussion of this issue in Chapter Four (p.66) it was hypothesised that the changing of task demands in experiments using the Walsh stimuli will result in changes in attentional strategy.

Experiment SIMWR (Similarity With Respect to Attention) used the same set of 15 stimuli as was used in experiment SIMS1. Experiment SIMWR was an attempt to quantify the effect of selective attention instructions on judgments of similarity between Walsh stimulus pairs. Although it cannot be assumed that selective attention instructions will necessarily result in selective attention, the rationale behind the use of the SIMWR results is as follows. If the results of a paradigm such as that of delayed similarity (Chapter 6) conform to those achieved under selective attention instructions, then this suggests, at least, that the changes in responding for the modified paradigm may be mediated by selective attention.

Inspection of the INDSCAL solutions for experiments SIMWR and SIMS1 and the correlations between the dimensions of the solution and the candidate features, indicated that three features appeared to account for the results. Regression models incorporating these three features as predictor variables were then fitted to the results of each participant in each of the two experiments. The scaling-regression approach used in this thesis was designed to identify salient features and their effects across different individuals in the absence of prior knowledge about the Walsh stimuli and their perceived structure. Analysis of variance designs should be of use in subsequent work with the Walsh stimuli, particularly in dealing with the problem of multicollinearity (high intercorrelation) of features which was evident in the regression analyses. However, since similarities are relative judgments, levels of independent variables will have to be constructed in terms of pairs of stimuli.

The regression models (4.2) and (4.3) on pp. 76-77 use psychological distance as the dependent variable. Gregson (1975, p. 192) used the following conversion from a distance in a scaling space D (a,b) between stimuli a, b, to obtain a similarity S (a,b):

$$S(a,b) = e^{-k \cdot D(a,b)}.$$

Luce (1961) developed a theory, based on choice-axiomatic considerations, where the relationship between similarity S and distance D was as follows:

$$D(a,b) = -\log S(a,b)$$

The results of SIMWR and SIMS1 were regressed on three features (complexity, component one, and sequency) using dissimilarities converted from similarities by means of a number of exponential and logarithmic transformations. The best fitting models (across a variety of parameter settings), as assessed by the multiple correlation coefficient, were found to give no better fit than was gained by using the relatively simple transformation (from similarity to dissimilarity):

$$D(a,b) = -S(a,b).$$

This does not imply that distance is simply negative similarity, but rather that the regression fitting procedure did not, in practice, distinguish between the transformations in terms of goodness of fit. Thus the regression analyses used in Chapter Four (results are presented on tables 4.1 and 4.2, p.78) had negative similarity as the dependent variable.

Two-dimensional sequency is analogous to the matrix grain measure which has been used as an information-theoretic measure of complexity (cf. Dorfman & McKenna, 1966). Thus tables 4.1 and 4.2 contrast two measures of complexity (one derived empirically, the other on theoretical grounds) with a statistically-based measure which cannot be given a psychological interpretation. It can be seen that single-predictor (using one of the two complexity-related features) models provide the best account of seven of the 11 SIMWR participants' results, whereas the corresponding proportion for SIMS1 was only three out of ten.

Another way to characterise the difference due to the selective attention instruction is to compare the percent variance accounted for when a single-predictor complexity model gave the best fit. Using R^2 as an estimate of variance accounted for, the total variance accounted for by single-predictor complexity models in SIMS1 is equivalent to the results of 1.2 out of the 10 participants, as compared with 3.9 out of the 11 participants in SIMWR. Thus the selective attention instruction produced a larger weighting towards complexity in making similarity judgments.

The delayed similarities experiments reported in Chapter Six were designed to change the nature of the processing used during a similarity judgment by requiring the retention of the first stimulus (or some representation of it) for different time intervals.

Experiments SIMD1 and SIMD8 elicited similarities between pairs of set one stimuli for successive, and eight second delayed, presentations, respectively. The mean square correlation coefficients for the SIMD1 (.22) and SIMD8 (.22) INDSCAL solutions were low compared with the corresponding measures for SIMS1 (.62) and SIMWR (.69) which suggests that the INDSCAL model was inappropriate for the results of these experiments. In contrast, the average correlation coefficient for the SIMD3 INDSCAL solution was .58. The reason for the bad fit of the INDSCAL solution to the SIMS1 results, but the relatively good fit of the SIMD3 INDSCAL solution, is unclear. The similarity of the INDSCAL model to regression models was shown on pp. 76-77, and consequently the poor fit of the INDSCAL model indicates that regression models would also be inappropriate for the SIMD1 and SIMD8 results.

It is possible that forgetting of the first stimulus under the delayed presentation condition may result in greater variability in responding. This would lead to poorer fit of the regression based models even where the participants were using a linear weighting rule in assigning similarity. Interpretation of the SIMD1 INDSCAL solution (p. 114) suggests that a single feature (or stable set of features) which is closely related to complexity, is used in making the delayed similarity comparisons.

Selective attention instructions were successful in making the participant attend to a particular feature. Delayed presentation appeared to have a similar effect for the set one stimuli, but not for the set three stimuli. The question of which strategies are induced by delayed similarity judgments cannot be answered properly without further study of the nature of the task itself. In particular, similarity models need to be specified which will account for the similarities elicited under delayed presentation conditions, and the role of forgetting of the first stimulus in the pair needs to be investigated.

The Role of Similarity

Cognitive theory and techniques from mathematical psychology have been combined in this thesis as part of an exploratory study of aspects of a theory of learning. Both cognitive, and mathematical, psychology have used the important notion of similarity. As a consequence similarity has been used in a number of different ways in this thesis.

In mathematical psychology, pairwise similarity data have been used as input to multidimensional scaling and clustering solutions which are often taken to be approximate representations of an underlying psychological space (p. 82). This role of similarity as part of a data analytic strategy is represented in figure 9.1. Here similarity is used to derive multidimensional scaling and clustering solutions which can then be used to estimate psychological space. MDS and clustering solutions may also be combined to give a configurational representation.

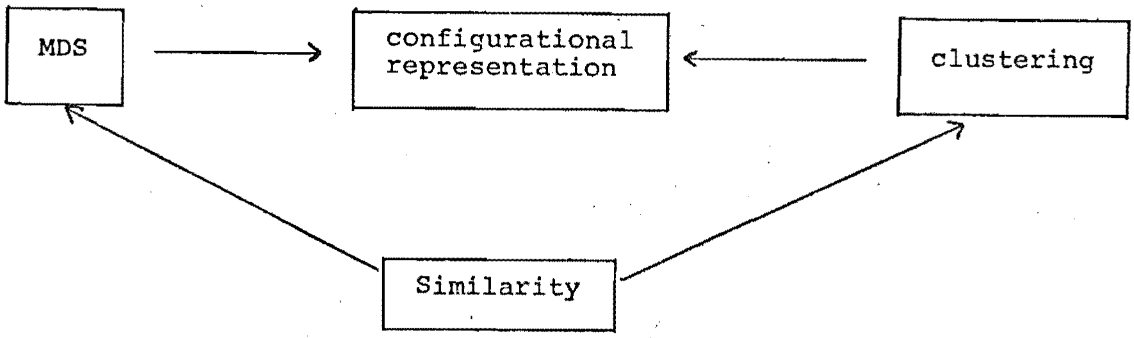


Figure 9.1 The role of similarity as a data collection method.

In cognitive psychology similarity is seen as a part of perceptual-cognitive processing as well as a phenomenon to be studied in its own right (pp. 82-86). Similarities are assumed to be produced by some feature comparison process which may also include the estimation of feature sizes. Thus the salience of features will be reflected in the similarity of objects which have particular values of those features.

Rosch (1978) has suggested that categories contain a single member (prototype) which is most representative of that category. There are a number of possible relationships between similarities and categories. A prototype can, for instance be interpreted as the member or members of the category with the highest average similarity to all members of the category. A category may also be viewed as being formed through a process of clustering based on similarities between objects in a feature space. This process is referred to as unsupervised learning in the machine pattern recognition literature and is briefly characterised by Andrews (1972). Thus similarity judgments can be viewed as part of the processes of category formation and category utilisation.

Salient features should be used in making both similarity judgments and categorisations. Thus the similarities between objects will define at least some of the discriminable features which differentiate those objects. This relationship is not generally reversable, however, as objects or concepts which differ markedly in terms of their component features may still be perceived as being similar through a process of analogy or metaphor (Ortony, 1979).

The role of similarity in cognitive theory is represented in figure 9.2. The feature space and the categories which an object belongs to are both part of the stored representation of the object.

A general representation of the role of similarity in this thesis is given in figure 9.3. Figures 9.1 and 9.2 are combined by the common notion of similarity. MDS may be linked with the feature space, as it was in this thesis where MDS solutions were an important step in the feature selection process.

Cluster analysis (on sorting data) was used to estimate the categories into which the Walsh stimuli were placed. In the present application a hierarchical clustering algorithm was used to partition the Walsh stimuli, but alternative approaches which allow a single stimulus to belong to a number of categories are also possible (e.g. Shepard & Arabie, 1979 and Appendix D of this thesis). Regression analysis was used to identify the features on which similarity judgments were based. The method of delayed similarities (to be discussed in the next section) was developed as a means of investigating the relationship between similarity and categorisation.

The upper half of figure 9.3 can be represented as data analytic uses of similarity, while the lower half represents similarity within a cognitive framework.

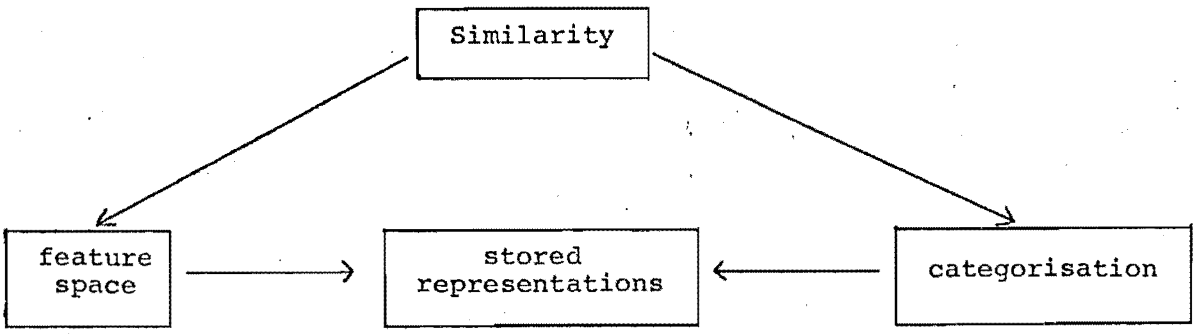


Figure 9.2 The role of similarity in cognitive theory.

Modified Similarity Paradigms

The method of collecting similarity judgments in the experiments reported in Chapters Four and Five may be modified in a number of ways. The first part of Chapter Six (pp. 100-111) considered some of the possible modifications, and attempted to derive a more general paradigm which would include a number of previous experimental tasks as special cases. A generalised similarities paradigm has been outlined where a variety of experimental tasks could be specified according to the way in which six parameters of a similarities task were set. The notion of generalised similarity was developed for two reasons. Firstly, to show the fundamental unity behind apparently disparate experimental tasks. Secondly, to provide a general introduction to the delayed similarity task as a special type of generalised similarity.

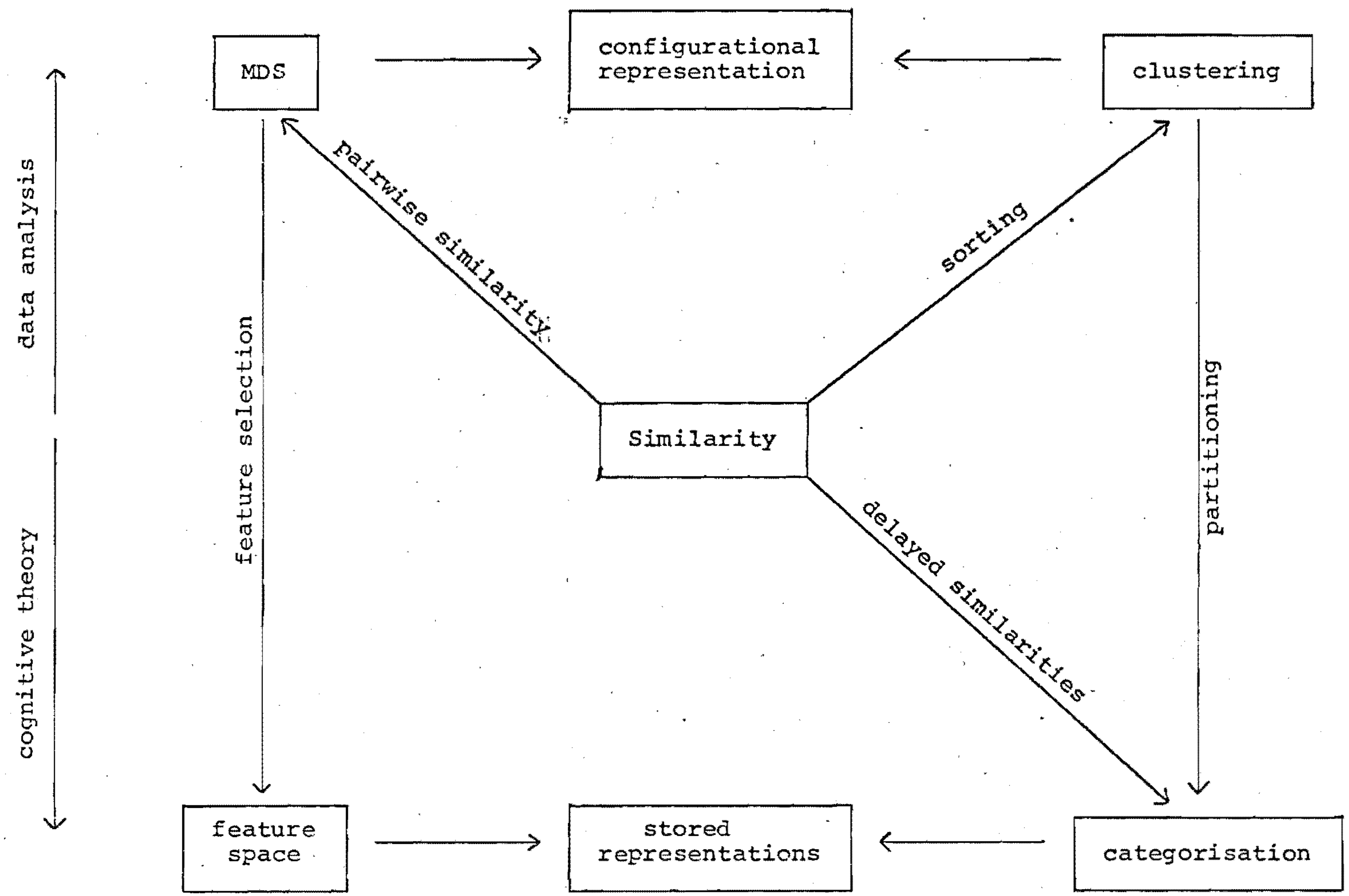


Figure 9.3 A schemantic representation of the use of similarity in terms of both data analysis and cognitive theory in this thesis.

Simultaneous visual presentation of stimulus pairs and memorial comparisons can be regarded as the two extremes on a continuum of memory involvement in similarity judgments. By varying the interstimulus interval in the delayed similarity task it should be possible to explore the three-way relationship between similarity, categorisation, and stored representations.

The studies reported in Chapter Six analysed delayed similarity judgments using the INDSCAL algorithm to get a general idea of the positioning of the stimuli in psychological space under delayed presentation conditions. The relationship between similarity and categorisation can be explored by deriving the category structure of the Walsh stimuli and then using the type of parameter settings for generalised similarity as were used by Posner (1978) with respect to same-different judgments.

The use of chronometric techniques in analysing the resulting data may prove difficult, however. Study of the relationship between similarity rating and reaction time (pp. 121-131) showed that this relationship differed markedly across individuals (figure 7.7, pp. 130-131). Even so, stimulus pairs which elicit comparatively long reaction times may prove to be of particular significance (e.g. tables 7.1 and 7.2, pp. 124-125).

Categorisation

The feature selection process identified five salient features, three of which were sequency-based, with the remaining two features (squareness and complexity) being closely related to sequency. Thus the Walsh stimuli were perceived and cognized in terms of the number of vertical and horizontal stripes of which they were made. Norman (cited on page 40) suggested that even things which we have never experienced before may already have a specific memory location reserved for them.[1] Such prior knowledge should result in consistent responding in the apparent absence of knowledge acquisition. The Walsh stimuli might then be seen as a subset of the set of possible checkerboard patterns and have memory locations reserved for them accordingly. This would mean that the categories into which the Walsh stimuli are grouped may already be defined before they are ever seen by an observer.

Four sorting experiments were carried out in Chapter Eight to investigate the perceived category structure of the Walsh stimuli. Four different sorting conditions were used to estimate the effect which small changes in the nature of the task had on the sorting results. The intercorrelations between the four proximity matrices derived from the sorting results (table 8.7, p. 152) were all greater than .6, while the proportion of matches (generally about 90%)

1. The issue of testing Norman's idea is not part of the present thesis although teaching a second language to an adult may provide relevant data.

between the incidence matrices was also high. Thus, the four sorting tasks seemed to indicate a single underlying category structure.

Table 8.6 (p. 151) gives the summary statistics for the proximities derived from the four sorting tasks. It can be seen that mean proximity has greatest in the Sort 1 condition, where a homogeneous black stimulus (Walsh 1) and a homogeneous white stimulus (an additional stimulus that was added to make 65 stimuli that were used in this experiment) were included in the stimulus set. The Sort 1 results produced a strong clustering structure. The contrast effect introduced by the homogeneous white stimulus may increase awareness of the checkerboard nature of the Walsh stimuli (p. 149) or it may increase the apparent proximity of the other stimulus pairs (p. 152). A visual interpretation of the Sort 1 clustering results is given on page 150 where membership in each of six clusters can be defined in terms of the number of vertical and horizontal stripes that each stimulus contains (pp. 149-150).

A fifth sorting task was run (p. 153) using only the 35 stimuli listed on page 70. While the pattern of results was similar to those for the Sort 1 task, there were marked individual differences in sorting strategy. Thus, while participants see the Walsh stimuli in terms of their vertical and horizontal stripes, they may differ in the way they partition the stimuli on the basis of that perceived structure. Walsh 8, for instance, may be seen as belonging either to a pile where column sequency is one, or to a pile where row sequency is eight.

Conclusions

The substantive findings of this thesis can be interpreted in two ways:

(1) As a study of the particular stimuli used (i.e. the Walsh stimuli).

(2) In terms of the development and testing of experimental tasks that may be used to derive an empirical theory of knowledge.

Psychological Properties of the Walsh Stimuli

Column sequency, row sequency, squareness, average grain, and complexity were identified as psychologically salient features of the Walsh stimuli.

It was not possible to assess the relative importance of these features because of their interdependence (multicollinearity) which resulted from the physical properties of the stimuli. Regression fitting of similarities elicited with simultaneously presented stimulus pairs suggested that complexity differences may be particularly important.

The Walsh stimuli were not evenly distributed throughout the feature space but formed clusters. The clustering analyses carried out in this thesis indicated the presence of three general clusters which correspond to three levels of sequency (p. 137).

Three clusters were also evident for the SIMS3 results (p. 98) and (to a lesser extent) the SIMS2 results (p. 97). The clustering of the sorting results reported in Chapter Eight yielded somewhat different results from the clusterings based on the similarity experiments. This may be due to the fact that the similarity clusterings were based on individual results while the sort clusterings were based on grouped results.

Six different scales of preference were identified across a sample of 44 participants. Preference scale one represented the results of 27 participants whose preferences increased monotonically with increasing complexity. The preference results of some of the participants appeared to be based on a feature which had not been quantified in Chapter Two. The ordering of the Walsh stimuli along this feature is shown in figure 3.9 (p. 59). Birkhoff (1933) suggested that preference was proportional to complexity and inversely proportional to a feature which he called 'order'. The present results show that preferences for the Walsh stimuli are based on complexity and a feature which may be characterised as order. The nature of this relationship between preference and these two features depends on the individual.

Under conditions of delayed presentation, similarities between pairs of Walsh stimuli appeared to be affected by forgetting of the first stimulus as well as by selective attention towards complexity in making the comparison.

Appendix A outlines some of the reasons for studying the psychological properties of the Walsh stimuli. The results reported in this thesis indicate that five features of the Walsh stimuli are psychologically salient. The preference structure of the Walsh stimuli has also been elaborated as a set of six single-peaked functions on a common bivariate feature space.

Knowledge Utilisation and Knowledge Representation

Chapters One and Four presented some of the relevant issues in cognitive theorising, but it proved difficult to devise experiments which could relate current cognitive theory to learning tasks. The type of theory outlined in figure 1.1 (p. 11), figure 1.2 (p. 12), figure 4.1 (p. 61), and figure 4.2 (p. 62) represent theories of knowledge utilisation in perception and cognition. The discussions in Chapters One (p. 4) and Four (pp. 60-63) show that there is considerable agreement between cognitive theorists on the issue of knowledge utilisation.

The problem of developing an adequate theory of knowledge acquisition is both difficult and unsolved. Knowledge acquisition involves a consideration of how knowledge is represented (pp. 7-8). Many theories of knowledge representation (often referred to by terms such as semantic memory) have been developed. Commonly agreed upon aspects of these theories have been summarised in Chapters One and Four. In practice, the representation of a given stimulus set has to be determined empircally. With visual stimuli the features and category structure are not intuitively obvious as they often appear to be with verbal stimuli (words). The apparent knowledge representation during acquisition will depend on the experimental tasks used, as well as the amount, and type, of previous experience of the stimuli that an observer has had (p. 8). Thus a method of experimentation and analysis was required which would give an appropriate estimate of cognitive representation during different stages of the learning process. In the present thesis similarities tasks and multidimensional scaling were used to estimate cognitive representations, along with sorting tasks and cluster analysis.

Summary

This thesis attempted to develop experimental tasks suitable for the study of knowledge acquisition. In particular, a closer relationship between theory, experimental task, and data analytic method was sought. The elaboration of the nature of stored representations for a particular stimulus set is necessary for investigating knowledge utilisation by means of generalised similarities tasks. Stored representations of the Walsh stimuli were assumed to consist of a representation of the objects in a feature space, combined with a representation of the objects within a system of categories.

The feature selection, and clustering, methods employed here can be used to identify the properties of stored representations for other stimulus sets. The generalised similarities paradigm which was developed in this thesis is a method for studying knowledge utilisation and representation during different stages of knowledge acquisition (learning). There are many ways in which this type of method can now be used, as can be seen by reference to the related work of Posner (1978), Posner & Rogers (1978), Schiffrin & Schneider (1977), and Schneider & Shiffrin (1977).

In developing a psychological theory of knowedge (an empirically-based epistemology), the stimuli used are particularly important. In spite of their apparent novelty, the Walsh stimuli appeared to be quickly assimilated in terms of their horizontal and vertical stripes.

It is clear that specific models of cognitive processes and structures are necessary if techniques developed in mathematical psychology are to be successfully applied. This thesis shows how the stored representation of a stimulus set may be derived.

The way in which stored representations change as knowledge is acquired can only be studied by using stimuli which (unlike the Walsh stimuli) do not elicit asymptotically stable stored representations when they are first presented.

It is possible that other forms of checkerboard stimuli may fulfill the above condition. Once a suitable set of stimuli has been found, the methods developed in this thesis (and extensions of those methods) may be used in quantitative studies of the acquisition of knowledge about visual patterns, by adults.

It remains to be seen how far an epistomology of visual patterns will take us towards the ultimate goal of a coherent psychological theory of knowledge.

ACKNOWLEDGEMENTS

I would like to thank my supervisor, Professor Gregson, for his generous advice, support and patience. I would also like to thank Paul Russell, who along with Professor Gregson has had a major influence on my general approach to psychology. The preparation of this thesis would not have been possible without the help of my family and friends. Especially warm thanks go to Winton Bell for his technical assistance, and to Don Smith, Anne Porter and Richard Tan.

REFERENCES

Afifi, A.A. and Azen, S.P. *Statistical Analysis: A Computer Oriented Approach*. New York: Academic Press, 1972.

Anderson, J.R. Arguments concerning representations for mental imagery. *Psychological Review*, 1978, *55*, 249-277.

Anderson, M.M. Information integration theory: A brief survey. In D.H. Krantz, R.D. Luce, R.C. Atkinson and P. Suppes, (Eds.), *Contemporary Developments in Mathematical Psychology*, Volume 2. San Fransisco: Freeman, 1974.

Andrews, H.C. *Introduction of Mathematical Techniques in Pattern Recognition*. New York: John Wiley and Sons, 1972.

Attneave, F. Dimensions of similarity. *American Journal of Psychology*, 1950, *63*, 516-556,

Attneave, F. and Arnoult, M.D. The quantitative study of shape and pattern perception. *Psychological Bulletin*, 1956, *53* 452-471.

Baddeley, A.D. The trouble with levels: A reexamination of Craik and Lockhart's framework for memory research. *Psychological Review*, 1978, *85*, 139-152.

Bamber, D. Reaction times and error rates for "same-different" judgements of multidimensional stimuli. *Perception and Psychophysics*, 1969, *6*, 160-174.

Bartlett, F.C. *Remembering: A Study in Experimental and Social Psychology*. Cambridge: University Press, London, 1932.

Bell, M.W., Gregson, R.A.M. and Kennedy, J.C. Universal Psychological Interface. *Research Report 29*. Department of Psychology and Sociology, University of Canterbury, New Zealand, 1978.

Berlyne, D.E. <u>Aesthetics and Psychobiology</u>. New York: Appleton-Century-Crofts, 1971.

Bezdek, J.C. Numerical taxonomy with fuzzy sets. <u>Journal of Mathematical Biology</u>, 1974, <u>1</u>, 57-74.

Birkhoff, G.D. Quelques elements mathematiques de l'Art, 1929.

Birkhoff, G.D. <u>Aesthetic Measure</u>. Cambridge, Mass.: Harvard University Press, 1933.

Blakemore, C. and Campbell, F.W. Adaptation to spatial stimuli. <u>Journal of Physiology</u>, 1969, <u>203</u>, 237-260.

Boorman, S.A. and Arabie, P. Structural measures and the method of sorting. In R.N. Shepard, A.K. Romney and S.B. Nerlove (Eds.), <u>Multidimensional Scaling, Volume 1: Theory</u>. New York: Seminar Press, 1972.

Broadbent, D.E. <u>Decision and Stress</u>. London: Academic Press, 1971.

Brown, D.R. and Owen, D.H. Metrics of visual form: Methodological dyspepsia. <u>Psychological Bulletin</u>, 1967, <u>68</u>, 243-259.

Bruner, J.S. On perceptual readiness. <u>Psychological Review</u>, 1957, <u>64</u>, 123-152.

Bruner, J.S. <u>Beyond the Information Given</u>. London: George Allen and Unwin, 1973.

Bruner, J.S., Goodman, J.J. and Austin, G.A. <u>A Study of Thinking</u>. New York: Wiley, 1956.

Carroll, J.D. and Arabie, P. <u>INDCLUS: A three-way approach to clustering</u>. Paper presented at Meeting of the Psychometric Society, Monterey, California, 1979.

Carroll, J.D. and Arabie, P. Multidimensional scaling. In M.R. Rosenzweig and L.W. Porter (Eds.), <u>Annual Review of Psychology</u>, 1980.

Carroll, J.D. and Chang, J.J. Analysis of individual differences in multidimensional scaling via an N-way generalisation of Eckart-Young decomposition. <u>Psychometrika</u>, 1970, <u>35</u>, 283-319.

Carroll, J.D. and Wish, M. Measuring preference and perception with multidimensional models. Bell Laboratories Record, 1972, 147-154.

Carroll, J.D. and Wish, M. Multidimensional perceptual models and measurement methods. In E.C. Carterette and M.P. Friedman (eds.), Handbook of Perception, Volume 2. New York: Academic Press, 1974.

Chignell, M. and Stacey, B.G. Practical problems associated with the use of cluster analysis. Psychological Reports, 1980, 46, 131-134.

Chipman, S.F. Complexity and structure in visual patterns. Journal of Experimental Psychology: General, 1977, 106, 269-301.

Coombs, C.H. A Theory of Data. New York: John Wiley and Sons, 1964.

Coombs, C.H. and Avrunin, G.S. Single-peaked functions and the theory of preference. Psychological Review, 1977, 84, 216-230.

Cormack, R.M. A review of classification. Journal of the Royal Statistical Society, (A) 1971, 134, 321-361.

Craik, F.I.M. Human memory. Annual Review of Psychology, 1979, 30, 63-102.

Craik, F.I.M. and Lockhart, R.S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Cupchik, G.C. and Berlyne, D.E. The perception of collative properties in visual stimuli. Scandanavian Journal of Psychology, 1979, 20, 93-104.

Davis, R.C. An evaluation and test of Birkhoff's aesthetic measure formula. Journal of General Psychology, 1936, 15, 231-240.

Dember, W.N. The Psychology of Perception. New York: Holt, Rinehart and Winston, 1960.

Dember, W.N. and Earl, R.W. Analysis of exploratory manipulatory, and curiosity behavior. Psychological Review, 1957, 64, 91-96.

Deutsch, J.A. and Deutsch, D. Attention: Some theoretical considerations. Psychological Review, 1963, 70, 80-90.

Dixon, W.I. (Ed.) B M D Biomedical Computer Programs. Berkeley: University of California Press, 1973.

Dixon, W.J. (Ed.) B M D P Biomedical Computer Programs. Berkeley: University of California Press, 1975.

Dorfman, D.D. and McKenna, H. Pattern preference as a function of pattern uncertainty. Canadian Journal of Psychology, 1966, 20, 143-153.

Dornic, S., Kunnapas, T. and Bratfisch, O. Subjective similarity as a function of exposure time and short term memory. Reprints from the Institute of Applied Psychology, the University of Stockholm, 1970, No. 9.

Draper, N. and Smith, H. Applied Regression Analysis. New York: John Wiley and Sons, 1966.

Egeth, H.W. Parallel vs. serial processes in multi-dimensional stimulus discrimination. Perception and Psychophysics, 1966, 1, 245-252.

Eisler, H. and Ekman, G.A. A mechanism of subjective similarity. Nordisk Psychologi, 1959, 11, 1-10.

Eysenck, H.J. and Castle, M. Training in art as a factor in the determination of preference judgments for polygons. British Journal of Psychology, 1970, 61, 65-81.

Fagot, R.F. Nested models of relative judgement applications to a similarity averaging model. Perception and Psychophysics, 1979, 26, 255-264.

Fischer, W. and Micko, H.C. More about metrics of subjective spaces and attention distributions. Journal of Mathematical Psychology, 1972, 9, 36-54.

Frith, C.D. The subjective properties of complex visual patterns. In E.L.S. Leeuwenberg and H.F.J.M. Buffart (Eds.), Formal Theories of Visual Perception. Chichester: Wiley, 1978.

Galanter, E. Textbook of Elementary Psychology. San Fransisco: Holden-Day, 1966.

Garner, W.R. The Processing of Information and Structure. New York: John Wiley and Sons, 1974.

Garner, W.R. Aspects of a stimulus: Features, dimensions, and configurations. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. New Jersey: Lawrence Erlbaum, 1978.

Gibson, E.J. Principles of Perceptual Learning and Development. New York: Appleton-Century-Crofts, 1969.

Gibson, J.J. The Senses Considered as Perceptual Systems. Boston: Houghton-Mifflin, 1966.

Gibson, J.J. The theory of affordances. In R. Shaw and J. Bransford (Eds.), Perceiving, Acting and Knowing. New Jersey: Lawrence Erlbaum, 1977.

Gilmartin, K.J., Newell, A. and Simon, H.A. A program modelling short-term memory under strategy control. In C.W. Cofer (Ed.), The Structure of Human Memory. San Fransisco: Freeman, 1976.

Glushko, R.J. Pattern goodness and redundancy revisited: Multidimensional scaling and hierarchical clustering analysis. Perception and Psychophysics, 1975, 17, 158-162.

Gregson, R.A.M. Aesthetic preference rankings for complex objects, with reference to the problem of hedonic summation. Unpublished Ph.D. Thesis, University of London, 1961.

Gregson, R.A.M. Some possible forms of lexicographic evaluation. Psychometrika, 1963, 28, 177-183.

Gregson, R.A.M. An aesthetic hedonic contrast paradox. Australian Journal of Psychology, 1968, 20, 225-232.

Gregson, R.A.M. Induced variants on quadratic similarity judgements. Research Project 16. Department of Psychology and Sociology, University of Canterbury, New Zealand, 1969.

Gregson, R.A.M. Quadratic similarities. British Journal of Mathematical and Statistical Psychology, 1972, 25 1-13.

Gregson, R.A.M. Psychometrics of Similarity. New York: Academic Press, 1975.

Gregson, R.A.M. A comparative evaluation of seven similarity models. British Journal of Mathematical and Statistical Psychology, 1976, 29, 139-156.

Gregson, R.A.M. A cognitive multivariate time series and its analysis. Acta Psychologica, 1978, 42, 277-291.

Gregson, R.A.M. Content and distance similarity models: A correction to Sjoberg. Scandanavian Journal of Psychology, 1979, 20, 127-128.

Gregson, R.A.M. Model evaluation via stochastic parameter convergence as on-line system identification. British Journal of Mathematical and Statistical Psychology, 1980, 33.

Haralick, R.M., Shanmugam, K. and Dinstein, I. Textural features for image classification. IEEE Trans. Syst., Man., Cyber., 1973, SMC-3, 610-621.

Hayek, F.A. The Sensory Order. Chicago: University of Chicago Press, 1954.

Helson, H. Adaptation-level Theory: An Experimental and Systematic Approach to Behavior. New York: Harper and Row, 1964.

Henderson, L. Pandemonium and visual search. Perception, 1978, 7, 97-104.

Hilgard, E.R. and Bower, G.H. Theories of Learning,3rd ed. New York: Appleton-Century-Crofts, 1966.

Hochberg, J.E. Perception, 2nd edition. Englewood Cliffs: Prentice-Hall, 1978.

Hochberg, J. and McAlister, E. A quantitative approach to figural "goodness". Journal of Experimental Psychology, 1953, 46, 361-364.

Homa, D., Rhoads, D. and Chambliss, D. Evaluation of conceptual structure. Journal of Experimental Psychology: Human Learning and Memory, 1979, 5, 11-23.

Hoosain, R. The process of negative or incongruent perceptual and combined perceptual/linguistic stimuli. British Journal of Psychology, 1977, 68, 245-252.

Hubert, L.J. and Levin, J.R. Evaluating object set partitions: Free sort analysis and some generalisations. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 459-470.

Hubert, L.J. Generalized concordance. Psychometrika, 1979, 44, 135-142.

Hyman, R. and Well, A. Judgements of similarity and spatial models. Perception and Psychophysics, 1967, 2, 233-248.

Hyman, R. and Well, A. Perceptual separability and spatial models. Perception and Psychophysics, 1968, 3, 161-165.

Jenkins, J.J. Four points to remember: A tetrahedral model of memory experiments. In L.S. Cermak and F.I.M. Craik (Eds.), Levels of Processing in Human Memory. New Jersey: Lawrence Erlbaum, 1979.

Johnson, S.C. Hierarchical clustering schemes. Psychometrika, 1967, 32, 241-254.

Joreskog, K.G. Structural analysis of covariance and correlation matrices. Psychometrika, 1978, 43, 443-477.

Kahneman, D. Attention and Effort. Englewood Cliffs: Prentice-Hall, 1973.

Keele, S.W. Attention and Human Performance. California: Goodyear Publishing Co., 1973.

Keller, F.S. and Schoenfeld, W.N. Principles of Psychology. New York: Appleton-Century-Crofts, 1950.

Kendall, M.G. Rank Correlation Methods. London: Griffin, 1970.

Kendall, M.G. and Stuart, A. The Advanced Theory of Statistics, Vol 3. London: Griffin, 1973.

Kirsner, K. and Sang, D.L. Visual persistence and code selection in short-term memory for letters. Journal of Experimental Psychology: Human Perception and Performance, 1979, 2, 260-276.

Kosslyn, S.M. Imagery and internal representation. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. New Jersey: Lawrence Erlbaum, 1978.

Kroll, N.E.A. and Parks, T.E. Interference with short-term visual memory produced by concurrent central processing. Journal of Experimental Psychology: Human Learning and Memory, 1978, 4, 111-120.

Krumhansl, C.L. Concerning the applicability of geometric models to similarity design: The interrelationship between similarity and spatial density. Psychological Review, 1978, 85, 445-463.

Kruskal, J.B. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 1964, 29, 1-27.

Kruskal, J.B. The relationship between multidimensional scaling and clustering. In J. van Ryzin (Ed.), Classification and Clustering. New York: Academic Press, 1977.

Kruskal, J.B. and Wish, M. Multidimensional Scaling. Beverley Hills, California: Sage, 1978.

Kunnapas, T. Acoustic perception and acoustic memory of letters. Acta Psychologica, 1968, 28, 161-170.

Lingoes, J.C. and Borg, I. A direct approach to individual differences scaling using increasingly complex transformations. Psychometrika, 1978, 43, 491-519.

Luce, R.D. A choice theory analysis of similarity judgments. Psychometrika, 1961, 26, 151-163.

Mace, W.M. James J. Gibson's strategy for perceiving: Ask not what's inside your head, but what your head's inside of. In R. Shaw and J. Bransford (Eds.), Perceiving, Acting and Knowing. New Jersey: Lawrence Erlbaum, 1977.

Mavrides, C.M. Selective attention and individual preferences in judgemental responses to multi-feature patterns. Psychonomic Science, 1970, 21(2), 67-68.

Micko, H.C. and Fischer, W. The metric of multi-dimensional psychological spaces as a function of the differential attention to subjective attributes. Journal of Mathematical Psychology, 1970, 71, 118-143.

Miller, G.A. A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 1969, 6, 169-191.

Miller, G.A., Galanter, E. and Pribram, K. Plans and the Structure of Behavior. New York: Holt, 1960.

Morris, C. Varieties of Human Value. Chicago: University of Chicago Press, 1956.

Morgan, B.J.T. Cluster analysis of two acoustic confusion matrices. Perception and Psychophysics, 1973, 13, 13-24.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Murch, G.M. Visual and Auditory Perception. Indianapolis: Bobbs-Merrill Co., 1973.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Neisser, U. Cognition and Reality. San Fransisco: W.H. Freeman, 1976.

Nelson, K. Cognitive development and the acquisition of concepts. In R.C. Anderson, R.J. Spiro and W.E. Montague (Eds.), Schooling and the Acquisition of Knowledge. New Jersey: Lawrence Erlbaum, 1977.

Nelson, K. Concept, word and sentence: Interrelationships in acquisition and development. Psychological Review, 1974, 81, 267-285.

Nickerson, R.S. "Same-different" response times with multi-attribute stimulus differences. Perceptual and Motor Skills, 1967, 24, 543-554.

Nickerson, R.S. and Adams, M.J. Long-term memory for a common object. Cognitive Psychology, 1979, 11, 287-307.

Norman, D.A. Toward a theory of memory and attention. Psychological Review, 1968, 75, 522-536.

Norman, D.A. Memory and Attention. New York: John Wiley and Sons, 1969.

Norman, D.A. and Bobrow, D.G. On the role of active memory processes in perception and cognition. In C.N. Cofer (Ed.), The Structure of Human Memory. San Fransisco: W.H. Freeman, 1976.

Oden, G.C. A fuzzy logical model of letter identification. Journal of Experimental Psychology, 1979, 5, 336-352.

Ortony, A. Beyond literal similarity. Psychological Review, 1979, 86, 161-180.

Palmer, S.E. Fundamental aspects of cognitive representation. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. New Jersey: Lawrence Erlbaum, 1978.

Peterson, L.R. Verbal learning and memory. Annual Review of Psychology, 1977, 28, 393-415.

Phillips, W.A. On the distinction between sensory storage and short-term visual memory. Perception and Psychophysics, 1974, 16, 283-290.

Phillips, W.A. and Baddeley, A.D. Reaction time and short-term visual memory. Psychonomic Science, 1971, 22, 73-74.

Posner, M.I. Chronometric Explorations of Mind. New Jersey: Lawrence Erlbaum, 1978.

Posner, M.I., Boies, S.J., Eichelman, W.H. and Taylor, R.L. Retention of visual and name codes of single letters. Journal of Experimental Psychology Monograph, 1969, 79, 1-16.

Posner, M.I. and Keele, S.W. Decay of visual information from a single letter. Science, 1967, 158, 137-139.

Posner, M.I. and Rodgers, M.G.K. Chronometric analysis of abstraction and recognition. In W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes, Volume 5, Human Information Processing. New Jersey: Lawrence Erlbaum, 1978.

Posner, M.I. and Snyder, C.R.R. Attention and cognitive control. In P.M.A. Rabbitt and S. Dornic (Eds.), Attention and Performance, V. London: Academic Press, 1975.

Proctor, R.W. Attention and modality-specific interference in visual short-term memory. Journal of Experimental Psychology: Human Learning, 1978, 4, 239-245.

Reed, S.K. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973.

Richardson, M.W. Multidimensional psychophysics. Psychological Bulletin, 1938, 35, 659-660.

Rosch, E. Human Categorization. In N. Warren (Ed.), Advances in Cross-Cultural Psychology, Volume 1. London: Academic Press, 1977.

Rosch, E. Principles of categorization. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. New Jersey: Lawrence Erlbaum, 1978.

Rosch, E. and Mervis, C.B. Family resemblances: Studies in the internal structure of categories. Cognitive Psychology, 1975, 7, 573-605.

Rosen, J. Symmetry Discovered: Concepts and Applications in Nature and Science. Cambridge: Cambridge University Press, 1975.

Rosler, F. Identifying interindividual judgement differences: Indscal or three-mode factor analysis. Multivariate Behavioral Research, 1979, 14, 145-167.

Rumelhart, D.E. and Ortony, A. The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro and W.E. Montague (Eds.), Schooling and the Acquisition of Knowledge. New Jersey: Lawrence Erlbaum, 1977.

Schank, R.C. and Abelson, R.P. Scripts, Plans, Goals and Understanding. New Jersey: Lawrence Erlbaum, 1977.

Schneider, W. and Shiffrin, R.M. Controlled and automatic human information processing, I: Detection, search and attention. Psychological Review, 1977, 84, 1-66.

Shepard, R.N. Attention and the metric structure of the stimulus space. Journal of Mathematical Psychology, 1964, 1, 54-87.

Shepard, R.N. Representation of structure in similarity data: Problems and prospects. Psychometrika, 1974, 39, 373-421.

Shepard, R.N. and Arabie, P. Additive clustering: Representations of similarities as combinations of discrete overlapping properties. Psychological Review, 1979, 86, 87-123.

Sherif, M., Taub, D. and Hovland, C.T. Assimilation and contrast effects of anchoring stimuli on judgements. Journal of Experimental Psychology, 1958, 55, 150-155.

Shiffrin, R.M. and Schneider, W. Controlled and automatic human information processing, II: Perceptual learning, automatic attending, and a general theory. Psychological Review, 1977, 84, 127-190.

Shipley, W.C., Coffin, J.I. and Hadsell, K.C. Affective distance and other factors determining reaction time in judgements of color preference. Journal of Experimental Psychology, 1945, 35, 206-215.

Sjoberg, L. A cognitive theory of similarity. Goteborg Psychological Reports, 1972, 2, No. 10.

Smets, G. Aesthetic Judgement and Arousal. Leuven, Belgium: Leuven University Press, 1973.

Smith, E.E., Shoben, E.J. and Rips, L.J. Structure and process in semantic memory: A featural model for semantic decisions. Psychological Review, 1974, 81, 214-241.

Sokal, R.R. and Sneath, P.M.A. Principles of Numerical Taxonomy. San Fransisco: Freeman, 1963.

Spence, I. Multidimensional scaling: An empirical and theoretical investigation. Unpublished Doctoral Thesis. University of Toronto, Canada, 1970.

Stotland, E. and Canon, L.K. Social Psychology: A Cognitive Approach. Philadelphia: Saunders, 1972.

Takane, Y., Young, F.W. and de Leeuw, J. Nonmetric individual differences multidimensional scaling: An alternating least squares method with optimal scaling features. Psychometrika, 1977, 42, 7-67.

Torgerson, W.S. Multidimensional scaling: I theory and method. Psychometrika, 1952, 17, 401-419.

Torgerson, W.S. Multidimensional scaling of similarity. Psychometrika, 1965, 30, 379-393.

Tucker, L.R. Relations between multidimensional scaling and three-mode factor analysis. Psychometrika, 1972, 37, 3-27.

Turvey, M.T. On peripheral and central processes in vision: Inferences from an information processing analysis of masking with patterned stimuli. Psychological Review, 1973, 80, 1-52.

Turvey, M.T. Contrasting orientations to the theory of visual information-processing. Psychological Review, 1977, 84, 67-88.

Turvey, M.T. Visual processing and short-term memory. In W.K. Estes (Ed.), Handbook of Learning and Cognitive Processes, Volume 5, Human Information Processing. New Jersey: Lawrence Erlbaum, 1978.

Tversky, A. Features of similarity. Psychological Review, 1977, 84, 327-352.

Tversky, A. and Gati, I. Studies of similarity. In E. Rosch and B.B. Lloyd (Eds.), Cognition and Categorization. New Jersey: Lawrence Erlbaum, 1978.

Valentine, C.W. The Experimental Psychology of Beauty. London: Methuen and Co., 1962.

Vernon, M.D. The functions of schemata in perceiving. Psychological Review, 1955, 62, 180-192.

Waterman, D.A. and Hayes-Roth, F. (Eds.) Pattern-Directed Inference Systems. New York: Academic Press, 1978.

Waugh, N.C. and Norman, D.A. Primary memory. Psychological Review, 1965, 72, 89-104.

West, E.M. and Bendig, A.W. Esthetic fatigue in ranking. American Journal of Psychology, 1954, 67, 285-287.

Whitfield, T.W.A. and Slatter, P.E. The effects of categorisation and prototypicality on aesthetic choice in a furniture selection task. British Journal of Psychology, 1979, 70, 65-75.

Wiener-Ehrlich, W.K. Dimensional and metric structures in multidimensional stimuli. Perception and Psychophysics, 1978, 24, 399-414.

Wish, M. and Carroll, J.D. Applications of individual differences scaling to studies of human perception and judgement. In E.C. Carterette and M.P. Friedman (Eds.), Handbook of Perception, Volume II: Psychological Judgement and Measurement. New York: Academic Press, 1974.

Woodworth, R.J. Experimental Psychology. New York: Holt, 1938.

Young, F.W. POLYCON: A program for multidimensional scaling, one-, two- or three-way data in additive difference, or multiplicative spaces. Behavioral Science, 1973, 18, 152-155.

Young, P.T. The role of hedonic processes in the organisation of behavior. Psychological Review, 1952, 59, 249-262.

Zadeh, L.A. Fuzzy sets. Information and Control, 1965, 8, 338-375.

Zusne, L. Visual Perception of Form. New York: Academic Press, 1970.

APPENDIX A

The present thesis has used a single set of stimuli which are shown in figure A.1. This appendix introduces the Walsh stimuli and shows, as an example of their usage, the implications that these stimuli have for models of similarity judgement.

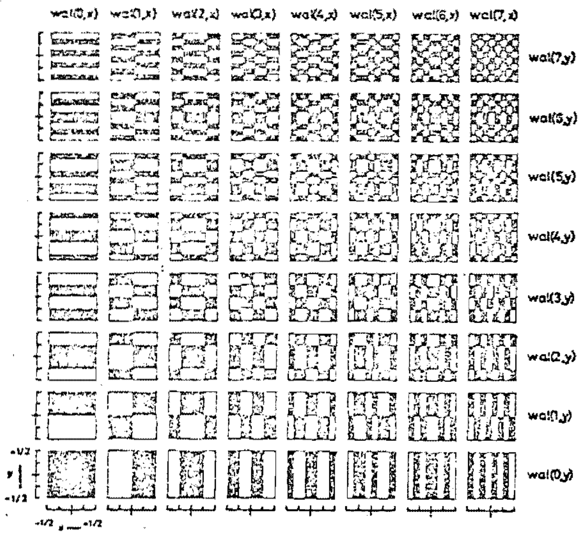
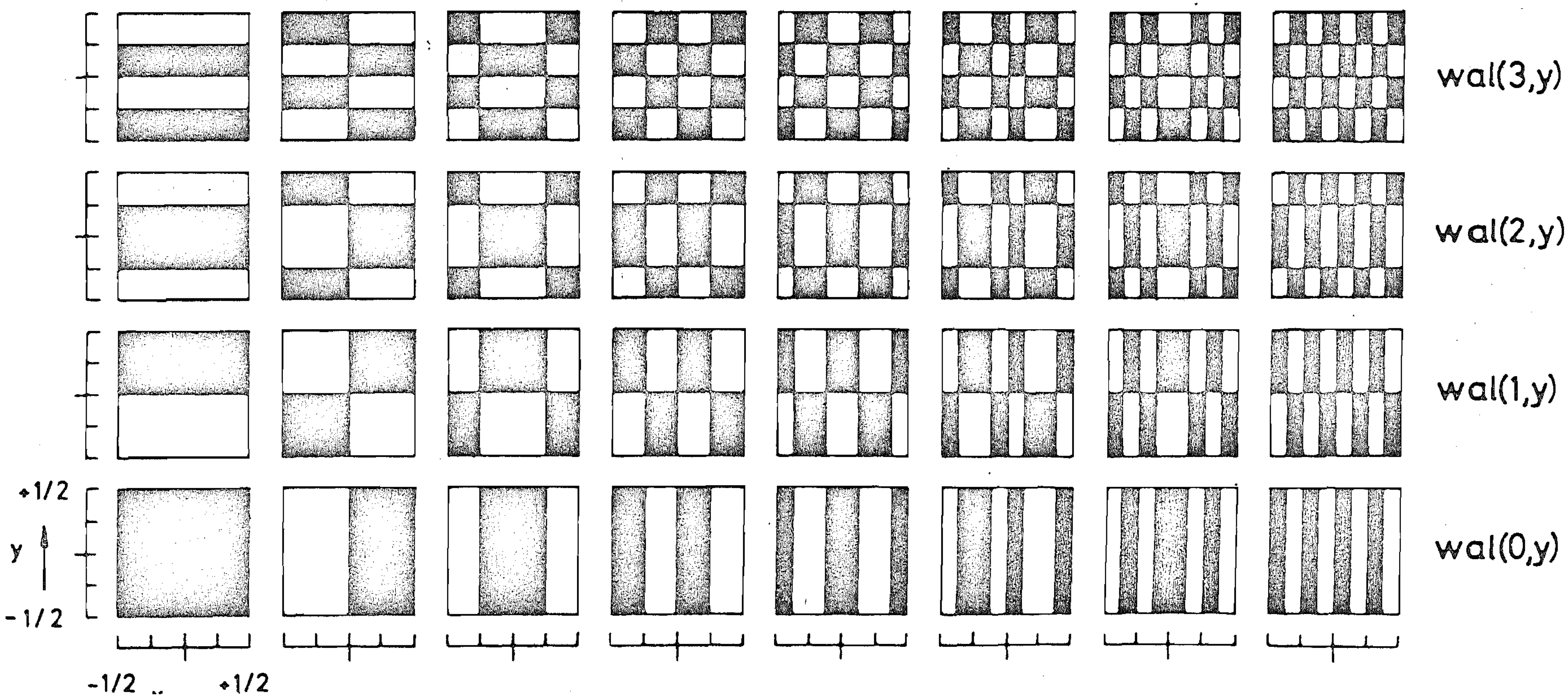


Figure A.1. The sequency-ordered matrix of Walsh stimuli

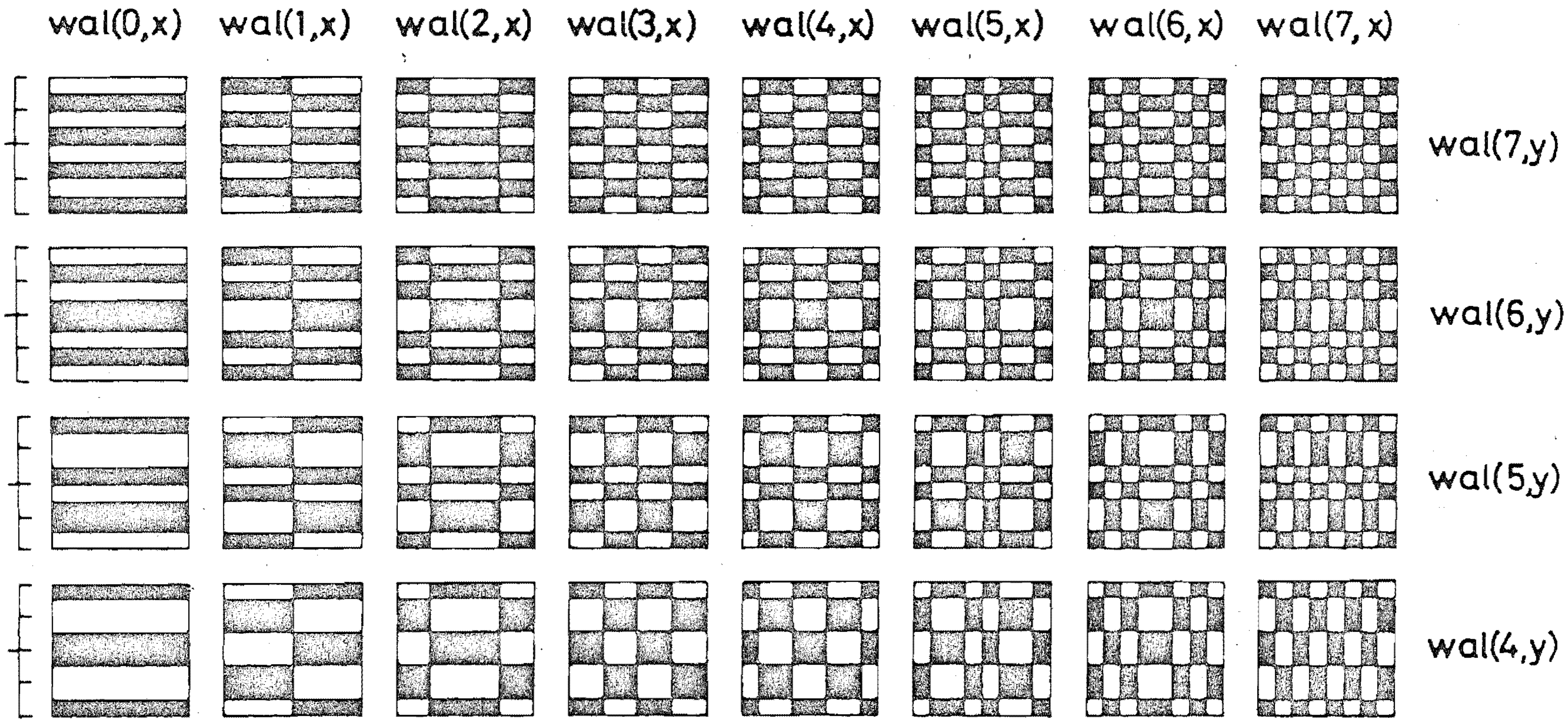
The history of visual form perception has been dominated by the search for adequate stimuli. Ideally, such stimuli should be both quantifiable, and ecologically valid (in the sense of Neisser, 1976).

Goldmeier's monograph (Goldmeier, 1972) is a good example of the Gestaltist approach to the problem. The Gestaltists sought to show general principles through the use of particular, well chosen, instances. This led to a hypothesis-testing approach which was inappropriate for the elucidation of quantifiable relationships. Perhaps as a reaction to the holistic approach of Goldmeier, Koffka, and others there followed a concerted drive towards quantification. Between 1956 and 1966 a number of largely information-theoretic constructions were proposed. These included the two methods for constructing random shapes (Attneave and Arnoult, 1956), the Ohio metric histoforms (Fitts and Leonard, 1957), metric polygons (Thurmond, 1966), and cellular polygons (Smith, 1964). These constructions have been summarised in Zusne (1970). Since 1970 there has been an increasing awareness of the cognitive component in visual form perception. The stimuli have been conceptualised less as a two-dimensional spatial grid and more as a collection of quantifiable dimensions or features. However the feature approach

[1] This appendix is derived from a paper presented to the 49th Congress of the Australian and New Zealand Association for the Advancement of Science; January 1979 at the Mathematical Psychology Symposium.



A close-up of figure A.1. Part I (bottom half)



A close-up of figure A.1. Part II (top half) has brought with it the realisation that the salience (weighting) of features may be expected to change with changes in the state of the organism . In many ways, this approach is analogous to previous hypotheses of motivated perception. Brown (1961) and Eriksen (1954) review the evidence for perceptual defence, but it is only recently that the issue of feature salience in perception has been looked at in a rigorous fashion. Barlett touched on the issue of stimulus saliences and response biases in Psychology when he said:

> "My own experiments show how, if consideration is confined to perceptual series, as the material given to be perceived increased in complexity, so its dominant characteristics, determined by orientation and attitude, may rapidly change. For the more complex the material, or its setting, the more varied is the play of interests and consequent attitudes which can be evoked".
>
> Bartlett (1932, p.193)

Tradeoff Considerations in Stimulus Design

1. The quantifiability of a stimulus set is inversely proportional to the ecological validity of that set. (This principle is illustrated in figure 1).

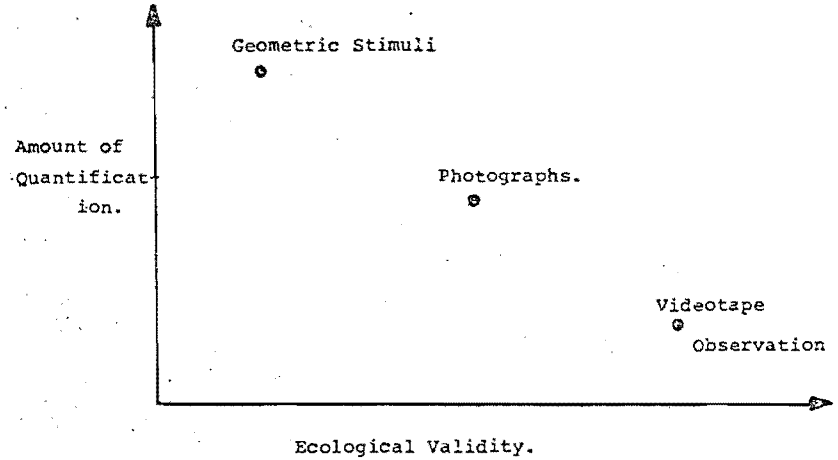


Fig 2: The tradeoff between quantification and ecological validity

2. The amount of response bias which can occur in judging a stimulus set is directly proportional to that set's ecological validity (i.e., the more complex and 'interesting' our stimuli become, the more we are involving the individual as a whole-biases, motivations and all - rather than just his perception of visual form).

Recent Stimulus Sets

Garner (1974) gives a good summary of the feature-dimension approach to form. Recent examples of stimuli embodying this approach are the triangle, circle, square stimuli of Gregson (1978), and the three ingeniously designed stimulus sets of Frith (1978), who used schematic faces, schematic imaginary animals, and histoform patterns.

There has also evolved a class of stimuli which are neither geometrical nor feature-dimensional. The one identifying feature of this class is that they all consist of an nxn grid of binary variables. The Walsh masks we are using are a special subset of the general set (with n = 8). While in one sense this is a return to 'two-dimensional spatial grid stimuli', the structure of such stimuli is as often featural as it is geometric. Chipman (1977) and Chipman et al, (1977) used a 6 x 6 grid to investigate perceived complexity, while Smets (1973) used three interesting sets (n = 8, n = 15, n = 30) in her study of aesthetic judgement and arousal.

Bias in Stimulus Construction

With the exception of the Walsh stimuli, every stimulus set discussed so far appears to have been constructed by a psychological researcher. Chipman discussed the problem of experimenter bias in the following manner:

> "...Ideally, an experiment of this kind (exploring pattern structure) uses a random sample of all simple (structured) stimuli possible within the stimulus domain. The rarity of highly structured patterns in randomly generated samples precludes that approach... Of course, the possibility of (experimenter) bias remains whenever the stimuli are experimenter designed".
>
> Chipman (1977, pp 272-3)

Psychologists have been faced with the conflicting goals of designing stimulus sets which are structured, but not structured so that they are more relevant to only one (i.e. they can be used in testing that theory) or a limited subset of the set of possible theories.

The two horns of this dilemma, plus the different paradigmatic stances taken (scaling versus similarity modelling and classification), appear to have largely determined the construction of stimuli in the field of visual form perception. Despite an abundance of stimuli, and factor analytic and multidimensional scaling analyses, we do not seem much closer to understanding the nature of visual form perception (C.f. Brown and Owen, 1967).

MODELLING PAIRWISE SIMILARITY JUDGEMENTS OF THE WALSH MASKS

We shall now demonstrate the quantifiable nature of the Walsh masks with particular reference to similarity modelling.

Gregson (1976) outlined seven models of similarity judgements for a stimulus pair comprising $\{x_i, x_2\} = X$ and $\{y_1, y_2\} = Y$, where the dimensions, i, are suffixed as one and two, and all x_i, y_i are real non-negative. Gregson chose these models because they were prototypical of a class of models that have been extensively used. They are by no means exhaustive, and in addition, they are mathematically interrelated as members of the Minkowski class of models, or the class of content models, or both. We shall show that such models are redundant for the Walsh masks (the models can be applied to the Walsh masks by expanding the number of dimensions to eight - one per row, or 64 - one for each binary variable) unless further assumptions are made. In particular, our notion of 'dimension' will have to be expanded to include secondarily-derived features of the Walsh masks.

Model 1

The first model is constructed by considering the sixty-four elements of the 8 x 8 grid to be perceptually distinguishable elements.

In this case, the pairwise similarity between two Walsh masks X, and Y, is given by

$$ {}^{1}\mathcal{S}(X, Y) = \sum_{i=1}^{64} k.\text{abs}\,(x_i + y_i) \qquad (1)$$

This model turns out to be redundant as in fact do all models based on the subtraction or addition of individual elements, because of the mutual orthogonality of the Walsh masks. This immediately eliminates three of the distance based models discussed by Gregson (M1, M2, M3).

Model 2

We shall now consider M4, the content model used by Gregson (1975, equation 5.462.3). This model is in fact a realisation of the set-theoretic content model (Gregson, 1975, equation 5.44.1), replacing the set operators according to the rules of Halmos (1950). In its primitive form, the model is given by

$$\mathcal{S}(X,Y) = \sum_{i=1}^{n} w_i \underline{m}\,(x_i \cap y_i) \Big/ \sum_{i=1}^{n} w_i \underline{m}\,(x_i \cup y_i) \qquad (2)$$

for two stimuli, X and Y, each with n quantifiable dimensions.

Footnote 1 In this expression, k is a constant which ensures that similarity between individual elements is one or zero. If the elements are coded as 1,-1 then k = ½, if 1,o then k = 1.

We apply this model to the Walsh stimuli in the following fashion.

$$\mathcal{S}(X, Y) \overset{\text{def}}{=} \sum_{i=1}^{8} W_i \sum_{j=1}^{8} \tfrac{1}{2} \text{ abs } (X_{ij} + Y_{ij})$$

The maximum possible intersection for the i^{th} row is given by

$$\max_i = \min (\#\text{ blacks in } X_i, \#\text{ blacks in } Y_i) + \min (\#\text{ whites in a}, \#\text{ whites in b}).$$

The value of $\max_i$ turns out to be a constant for all pairs drawn from 56 of the Walsh masks (this affects about 75% of the possible pairs).

The full version of model 2 is as follows:

$$\mathcal{S}_2 (X,Y) = \frac{\sum_{i=1}^{8} W_i \sum_{j=1}^{8} \tfrac{1}{2} \text{ abs } (X_{ij} + Y_{ij})}{\sum_{i=1}^{8} W_i \text{ Max}_i} \qquad (3)$$

Both the denominator and the numerator of this model are redundant (the numerator fails because of the orthogonality of the Walsh functions). As a check on our reasoning, the model was tested for redundancy using computer simulation with a variety of parameter settings. The results given in Table 1 are as expected. Regardless of the weights chosen, the model predicted similarities of .5 for about 90% of the possible pairings. The first six simulations show that the model is relatively insensitive to the high- versus low- bias distinction. The results on simulation seven showed similar insensitivity to a wider range of parameter values.

TABLE 1

Simulation Results for Model 2

Simulation number	ROW BIASES									
	R1	R2	R3	R4	R5	R6	R7	R8	$S_{xy} \triangleq .5$	# Other distinct variables
1.	.5	.1	.1	.1	.1	.1	.1	.5	100	0
2	.5	.1	.1	.1	.1	.1	1.0	.5	94	1
3	.5	.1	.1	.1	.1	1.0	.1	.5	94	1
4	.5	.1	.1	.1	.1	1.0	1.0	.5	94	1
5	.5	.1	.1	.1	1.0	.1	.1	.5	94	1
6	.5	.1	.1	.1	1.0	1.0	1.0	.5	94	1
7	.5	.2	4	.8	.1	1.6	3.2	.5	94	1

Redundancy and the Walsh masks

In our brief survey of the models covered by Gregson (1976) we have so far found that five of the seven, i.e. M1, M2, M3, M4, and M5 are redundant with respect to the Walsh masks. M6 and M7 are also redundant in this case as they are based on the supremum metric which, as it is a distance measure, has already been shown to be inadequate due to the orthogonality property of the Walsh masks.

We have shown that seven of the most commonly used scaling models are inadequate with respect to our set of stimuli. Yet it is an observed fact that participants find the task of judging pairwise similarity of Walsh masks meaningful.[1] Our results imply that much of the previous similarity modelling has been a non-too-successful exercise in model-fitting for very specialised stimulus sets rather than the embodiment of a general process of similarity judgement in a quantified algebraic form.

A Possible Model

It appears that there is little in the previous similarity modelling literature that is relevant to the Walsh masks. Our third model attempts to use the special features of the Walsh masks - along with the element by element comparisons already used. The two special features we shall look at are block structure and sequency.

[1] Note the results obtained in Chapters four to six of this thesis

Block Structure

Each of the 64 Walsh masks can be conceptualised as a 4 x 4 matrix of black and white squares (we call such a 4 x 4 matrix a block) which is then expanded into an 8 x 8 matrix. The expansion is done by creating three other blocks (multiplying the original block by -1 if necessary) and placing them in a block schema as shown in figure 3

Block 1	Block 2
Block 3	Block 4

Fig. 3 a The Block Schema for a Walsh mask

S_1	S_2
S_3	S_4

Fig. 3 b The Block Structure for a Walsh mask

We shall call the original 4 x 4 matrix (i.e. Block 1) B. Then every other block is equivalent to S_iB where S_i is a scalar multiplier which can take the values +1 or -1 depending on the value of i and the particular Walsh mask referred to.

The block structure can be used to quantify the symmetries exhibited by the masks. It turns out that there are four possible block structures.

The two-dimensional block structures arise naturally from the corresponding one-dimensional structures. In one dimension, a Walsh function has one of the following schematic forms about its midpoint.

1. 1 1   2. 1 -1

If we denote two one-dimensional Walsh masks by W_i and W_j then the derived two-dimensional Walsh mask W_{ij}, say, is simply the matrix multiple, that is,

$$W_{ij} = W_i \cdot W_j^T$$

In analogous fashion, if we let the one-dimensional structure be $M(W_i)$ and the two-dimensional structure be $B(W_{ij})$, then we have

$$B(W_{ij}) = M(W_i) \cdot M(W_j)^T$$ (where T symbolises matrix transposition).

We show this process schematically in fig. 4.

1.	(1 1)	2.	(1 -1)	3.	(1 1)	4.	(1 -1)
1	1 1	1	1 -1	1	1 1	1	1 -1
1	1 1	1	1 -1	-1	-1 -1	-1	-1 1

Fig. 4 Evolution of the four types of two-dimensional block structure in the Walsh masks.

Table 2 gives the two-dimensional block structures for the 64 Walsh masks. It is laid out to show each two-dimensional Walsh mask in terms of its derivation from the two one-dimensional Walsh masks which provide its row and column. The numbers refer to the labels used in fig. 3.

	W_1	W_2	W_3	W_4	W_5	W_6	W_7	W_8
W_8	1	2	2	1	1	2	2	1
W_7	3	4	4	3	3	4	4	3
W_6	3	4	4	3	3	4	4	3
W_5	1	2	2	1	1	2	2	1
W_4	1	2	2	1	1	2	2	1
W_3	3	4	4	3	3	4	4	3
W_2	3	4	4	3	3	4	4	3
W_1	1	2	2	1	1	2	2	1

TABLE 2 Block structures for the two-dimensional Walsh masks.

Relative Block Structure

For any two Walsh stimuli, there are sixteen possible combinations of block structure given by the Cartesian product

(1, 2, 3, 4) x (1, 2, 3, 4).

We represent the block structure of X as B_x and the perceived relative block structure of X and Y as $\Psi(B_x, B_y)$ which is a relation on the interval [0, 1] . We now introduce the three following axioms for perceived relative block structure which have the desirable effect of reducing the number of parameters to be separately estimated.

Axiom 1. If $B_x = B_y$ then $\Psi(B_x, B_y) = 1$.

Axiom 2. $\Psi(B_x, B_y) = \Psi(B_y, B_x)$

Axiom 3. If $B_x \neq B_y$ then $(B_x, B_y) = P_{xy}$,
where $x \in \{1,2,3,4\}$, $y \in \{1,2,3,4,\}$.

where P_{xy} is an hypothesised perception of relative block structure which will vary from individual to individual. The formal parallel of these axioms to the axioms for a semimetric is not entirely fortuitous. The six values of P_{xy} which need to be determined for each individual are

P_{12}, P_{13}, P_{14}, P_{23}, P_{24}, and P_{34}.

Sequency

The sequency of a one-dimensional Walsh mask is defined to be half the number of zero crossings. In similar fashion, we define the sequency of a two-dimensional mask to be the number of patches of black or white making up the mask. The sequency of a two-dimensional Walsh mask corresponds to the product of the sequencies of its two component one-dimensional Walsh masks.

Relative Sequency

We shall denote the two-dimensional sequency of a mask X as Z_x.

The relative sequency $(Z_{(x,y)})$ between two masks X and Y is defined to be

$$Z_{(x,y)} \overset{\text{def}}{=} \min(Z_x, Z_y) \,/\, \max(Z_x, Z_y)$$

The corresponding psychologically perceived relative sequency ($\phi_{(x,y)}$, say), is then defined as

$$\phi_{(x,y)} \overset{\text{def}}{=} Z_{(x,y)}^{\,m} \qquad \text{for } 0 \leq m \leq 1.$$

If the exponent m equalled zero for a participant, this would indicate that he was not utilising relative sequency in his judgements.

For the present we have constrained the parameter m to be less than or equal to one, although it could conceivably be greater than one.

Block Similarity

We return to the element by element comparisons attempted previously as the final component in our model.

The Block Similarity is as in (1), but with the multiplier k removed and the matrix reduced to four rows and four columns:

$$\mathcal{S}_{block}(x, y) \overset{def}{=} \sum_{i=1}^{4} \sum_{j=1}^{4} \text{abs } (x_i + y_i).$$

We now have three feature-type components from which we can generate a non-metric sequency-based similarity model. Since the features are not independent we posit a multiplicative model of the following form.

A Proposed Model for Walsh pairwise Similarities

The Walsh pairwise similarity between X and Y is defined to be

$$\mathcal{S}_{w}(X,Y) \overset{def}{=} Z_{(x,\, y)}^{\;m} \cdot P_{xy} \cdot \mathcal{S}_{block}(X,Y)$$

Initially, this model involves a total of eight parameters. These are: m - the exponent relating psychologically perceived relative sequency to relative sequency.

$P_{x,y}$ - the six relative block structure parameters.

Implications for Similarity Modelling in General

The inability of a number of extensively used similarity models to cope with an unusual, but not invalid, stimulus set has wide implications for the whole field of similarity modelling. The usefulness of certain of the models based on Minkowski distances has already been questioned in the literature. For instance Gregson (1976) found M1, M2 and M3 to be unsatisfactory in his experiments. Another major class of models are the content models. These models are under-specified in the sense that there is generally no single obvious way of converting the set-theoretic operations of intersection and union into a useable form. In many cases the content model actually used to fit the data is closely related to the distance models and suffers from similar weaknesses, as was the case with the redundancy of the content model we used with our Walsh masks. Our results with the Walsh masks cast serious doubts on the generalisability of similarity models which have been constructed with reference to previous stimulus sets. One promising way of reconceptualising a stimulus set may involve the notion of transformation structure. If one looks at the 64 Walsh masks set out in a sequency ordered 8 x 8 matrix, one can see that opposite elements (with respect to the diagonal going from upper right to lower left) are identical except for a 90 degree rotation. When we also take into account the mathematical interdependence of the Walsh masks, we have good reason for suspecting that a weak transformational structure (at the very least) will be perceived by our participants.

Imai (1976, Handel and Imai, 1972) created a model based on inter-configurational transformations and an additive measure of the size of the variant features under transformation. This model produced meaningful solutions. Imai viewed the similarity judgement of patterns as a part of more general pattern cognition which should be looked upon as a basically dynamic process, typically represented by cognitive transformations. The recent work of Gregson (1978a, 1979, personal communication) has involved the study of similarity judgements as a dynamic process with a high degree of flexibility in the models as they attempt to track the participants' behaviour over time. This is an heuristic response to the mounting evidence that we need to broaden our conceptualisation of similarity and consequently the scope of our modelling.

Conclusion

The preliminary results given above have challenged some of the preconceived ideas in visual form perception and similarity modelling. Ideally, our theories should be able to handle stimulus sets from all parts of the continuum of ecological validity. Our results imply that existing theories may be ill-equipped to explain the perception and judgement of Walsh masks. We can hardly expect them to perform any better at the other end of the ecological validity continuum (e.g. with actual facial expressions), when they have been formulated with respect to a limited set of geometric forms and a few colours.

One promising approach may be to build a general theory of the perception of gray-coded n x n matrices. Presently available techniques such as the digitisation of photographs with flying-spot scanners could produce sets of stimuli of uniform quantifiability. Furthermore, these stimuli would range from the Walsh stimuli as far as black and white photographs of real-world situations in terms of their relevance to the everyday human environment.

<u>References</u>

Attneave, F. and Arnoult, M.D. The Quantitive Study of Shape and Pattern Perception. <u>Psychol. Bull.</u> <u>53</u>, 452 - 71 (1956).

Bartlett, F.C. <u>Remembering: A study in Experimental and Social Psychology</u>. Cambridge University Press, London (1932).

Brown, L.P. Conceptions of Perceptual Defence. <u>Br. J. Psychol. Monogr. Supp.</u>, <u>35</u>, Cambridge University Press (1961).

Brown, D.R. and Owen, D.H. The metrics of Visual form: Methodological Dyspepsia. <u>Psychol. Bull.</u>, 1967, <u>68</u>, 243-59 (1967).

Chipman, S.F. Complexity and Structure in Visual Patterns. <u>Journal of Experimental Psychology</u>: General, <u>106</u>, 269-301 (1977).

Chipman, S.F., Mendelson, M.J. and Waldner, D. Sensitivity to Visual Structure. Paper presented to the Society for Research in Child Development, New Orleans, March 17 (1977).

Ericksen, C.W. The case for perceptual defence. <u>Psychol. Rev.</u>, <u>61</u>, 175-82 (1954).

Fitts, P.M. and Leonard, J.A. Stimulus correlates of visual pattern recognition: A probability approach. Columbus, Ohio: Ohio State University (1957).

Frith, C.D. and Frith, U. Feature selection and Classification: A Developmental Study. <u>J. Experimental Child Psychol.</u>, <u>25</u>, 413-28 (1978).

Garner, W.R. The Processing of Information and Structure. Potomac, Maryland: Erlbaum (1974).

Goldmeier, E. Similarity in visually perceived forms. <u>Psychological Issues</u>, <u>8</u> (1), Monograph 29 (1972).

Gregson, R.A.M. <u>Psychometrics of Similarity</u>. New York: Academic Press (1975).

Gregson, R.A.M. A comparative evaluation of seven similarity models. <u>Br. J. Math. Statist. Psychol.</u>, <u>29</u>, 139-56 (1976).

Gregson, R.A.M. A cognitive multivariate time series and its analysis. <u>Acta Psychologica</u>, <u>42</u>, 277-91 (1978).

Gregson, R.A.M. Model Evaluation via stochastic parameter convergence in quasi-closed feedback judgement sequences. Paper read at the Mathematical Psychology symposium of the annual meeting for the Australia and New Zealand Association for the advancement of Science, January (1979).

Halmos, P.R. <u>Measure Theory</u>. New York: Van Nostrand (1950).

Harmuth, H.H. <u>Sequency theory: foundations and applications.</u> New York: Academic Press (1977).

Imai, S. Pattern Similarity and Cognitive Transformation. Hokkaido Report of Psychology, HRP-12-795-18 (1976).

Handel, S. and Imai, S. The free classification of analyzable and unanalyzable stimuli. Perception and Psychophysics, 12, 108-116 (1972).

Neisser, U. Cognition and Reality. San Francisco: W.H. Freeman and Company (1976).

Smets, G. Aesthetic Judgement and Arousal. Leuven University Press (1973).

Smith, J.P. The effect of figural shape on the perception of area. Unpublished doctoral dissertation, Fordham Univ. Diss. Abstr. 1964, 25 (b), 3712 (1964).

Thurmond, J.B. Extensions of an information deductive analysis of form. Paper read at a meeting of the Southern Society for Philosophy and Psychology, New Orleans, April (1966).

Walsh, J.L. A closed set of normal orthogonal functions. Am. J. Math., 45, 5-24 (1923).

Zusne, L. Visual Perception of Form. New York: Academic Press (1970).

APPENDIX B

This Appendix can be divided into two parts: The first part outlines some of the mathematical properties of Walsh functions while the second part supplements the information on the derivation of features of the Walsh stimuli which was given in Chapter two of this thesis.

The following sections will attempt to make the mathematical properties underlying the Walsh stimuli explicit.

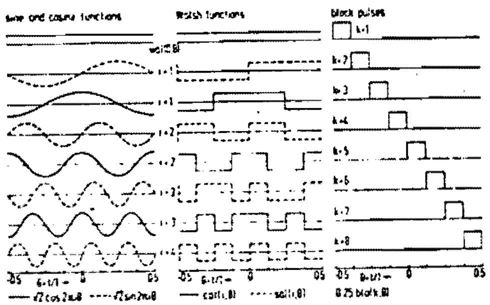


Figure 1. Sine-Cosine functions, Walsh functions, and Block pulses.

ORTHONORMALITY

A system $f(j,x)$ of real and almost everywhere non-vanishing functions $f(0,x)$, $f(1,x)$ .... is called orthogonal in the interval $x_0 \leq x \leq x_1$ if the following condition holds true ;

$$\int_{x_0}^{x_1} f(j,x)\cdot f(k,x)\,dx = X_j\,\delta_{jk}$$

$$\text{where } \delta_{jk} = \begin{cases} 1, & j=k \\ 0, & j\neq k \end{cases}$$

The functions are called orthogonal and normalised (orthonormal) if the constant X_j is equal to 1. Figure 1 shows three examples of orthogonal functions. (after Harmuth, 1977 p.3)

(The next section refers specifically to this figure which may be found in the rear).

The independent variable is the normalised time $\theta = t/T$. Both the sinusoidal and Walsh functions of figure 1 are orthonormal in the interval $-\frac{1}{2} \leq \theta < \frac{1}{2}$. In analogous fashion to the sinusoidal functions, the Walsh functions may be divided into even functions, $f_c(i,\theta)$ (cf the cosine functions), odd functions, $f_s(i,\theta)$ (cf. the sine functions), and the constant 1 ($Wal(0,\theta)$) .

Figure 2 (after Harmuth, 1977, p.21), shows the orthonormal system of Walsh elements, consisting of a constant $wal(0,\theta)$, even functions $cal(i,\theta)$, and odd functions $sal(i,\theta)$. (Note that $sal(i,\theta) = wal(2i-1,\theta)$ and $cal(i,\theta) = wal(2i,\theta)$.

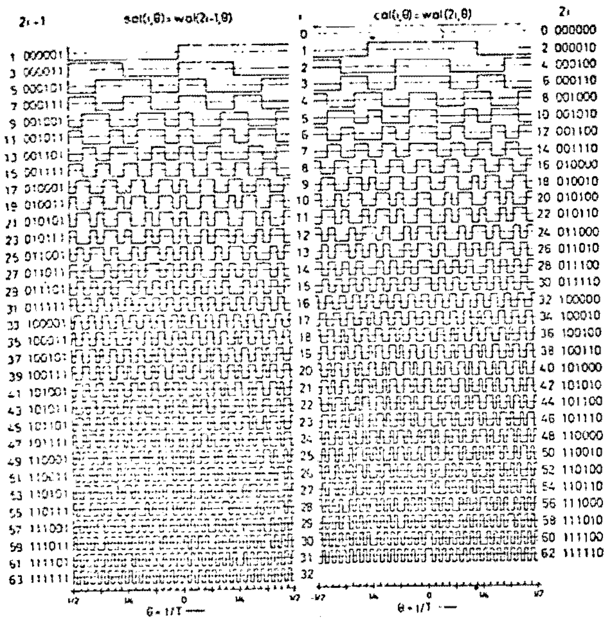


Figure 2. The orthonormal system of Walsh elements.

The fuctions jump back and forth between +1 and -1. The Walsh elements may be considered in blocks of 2^n with each block being describable by an n-digit binart code. (This binary code is shown on the outer sides of figure 2) .

For instance, the first 8 elements may be coded as 000, 001, 010, 011, 100, 101, 110, and 111. These 8 elements form the complete set of Walsh functions using 8 sample points. With 7 binary digits one would have 2^7 = 128 sample points and 128 Walsh elements. The term <u>element</u> is used to emphasise that a function is defined on a finite interval only and is undefined outside that interval.

<u>PROOF OF ORTHONORMALITY.</u>

1/. The integral of the product of any two Walsh functions is equal to zero. (This can be easily verified for any given pair of Walsh functions).

2/. A function multiplied with itself yields the products $(-1)\cdot(-1)$ or $(1)\cdot(1)$. Hence these products have the value 1 in the whole interval and their integral over that interval is 1.(Since the interval is of unit length).

COMPLETENESS

The set of orthogonal functions $f(j,\theta)$ is complete if the only function $h(\theta)$ that satisfies

$$\int h(\theta)\ f(j,\theta)d\theta = 0\ ,$$

for all j, is a null function, i.e,

$$\int h^2(\theta)d\theta = 0\ .$$

The Walsh functions are a complete set of orthogonal functions (Walsh, 1923) .

DEFINING THE WALSH FUNCTIONS.

The functions $wal(j,\theta)$ may be defined by the following difference equation ;

$$wal(2j+p,0) = (-1)^{[j/2]+p}\ wal\ j,2(0+\tfrac{1}{2})$$
$$+(-1)^{j\ +p}\ \ wal\ j,2(0-\tfrac{1}{2})$$

with p=0 or1, j=0,1,2,....
and $[j/2]$ means the largest integer smaller than or equal to j/2.

We can now derive the rest of the series from

$$Wal(0,\theta) = \begin{cases} 1, & \tfrac{1}{2} \leq \theta < \tfrac{1}{2} \\ 0, & \theta < -\tfrac{1}{2},\ \theta \leq \tfrac{1}{2}\ . \end{cases}$$

This difference equation can be given a clear visual meaning. Consider $Wal(j,\theta)$. The function $wal(j,2\theta)$ has the same shape, but is squeezed into the interval $-\tfrac{1}{2} \leq \theta < \tfrac{1}{2}$.

$Wal\ \ j,2(\theta+\tfrac{1}{2})$ is obtained by shifting $wal\ (j,2\theta)$ to the left into the interval $-\tfrac{1}{2} \leq \theta < 0$, and $Wal\ \ j,2(0-\tfrac{1}{2})$ is obtained by shifting $wal(j,2\theta)$ to the right into the interval $0 \leq \theta < +\tfrac{1}{2}$.

It can be seen from figure 2 that $wal(1,\theta)$ is obtained from $Wal(0,\theta)$ by squeezing it to half its width, multiplying the function that is shifted to the left by -1, and the function that is shifted to the right by +1.

MATHEMATICAL PROPERTIES OF THE WALSH FUNCTIONS.

1/. The product of two Walsh functions yields another Walsh function;

$$wal(h,\theta).wal(k,\theta) = wal(r,\theta)$$
$$r = h\oplus k$$

where $\oplus$ implies modulo 2 summation (no carry).
e.g. $wal(1,\theta)\ wal(3,\theta) = wal(2,\theta)$
since $01 \oplus 11 = 10$.

2/. A Walsh function multiplied by itself yields $wal(0,\theta)$ since only the products (+1) (+1) and (-1) (-1) occur. i.e. $i\oplus i = 0$

3/. A Walsh function multiplied by $wal(0,\theta)$ remains unchanged;

$$wal(j,\theta)\ wal(0,\theta) = wal(j,\theta)$$
since $j\oplus 0 = j$.

4/. The multiplication of Walsh functions is associative since only products of +1 and -1 occur.

i.e. $$[wal(h,\theta)wal(j,\theta)]\ wal(k,\theta) = wal(h,\theta)\ [wal(j,\theta)wal(k,\theta)]\ .$$

Walsh functions form a group with respect to multiplication because the previous four properties exist.

The group of Walsh functions is an Abelian group and is isomorphic to the discrete dyadic group.

THE USE OF WALSH FUNCTIONS IN FUNCTIONAL ANALYSIS

A function f(t) can be approximated by a set of Walsh functions as follows;

$$F(\theta) = \sum_{n=0}^{\infty} A(n)\ \mathrm{wal}(n,\theta)$$

or, in terms of cal and sal functions

$$F(\theta) = A(0)\ \mathrm{wal}(0,\theta) + \sum_{n=1}^{\infty}\left[A_c(n)\,\mathrm{cal}(n,\theta) + A_s(n)\,\mathrm{sal}(n,\theta)\right]$$

where $A(0) = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(\theta)\,\mathrm{wal}(0,\theta)\,d\theta = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(\theta)\,d\theta$

$$A_c(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(\theta)\,\mathrm{cal}(n,\theta)\,d\theta$$

$$A_s(n) = \int_{-\frac{1}{2}}^{\frac{1}{2}} F(\theta)\,\mathrm{sal}(n,\theta)\,d\theta$$

The time required to obtain the Fourier transforms may be drastically reduced by means of a method known as the fast Fourier transform. There is also a fast Walsh-Fourier transform (Welch 1966) which only requires $2^n * n$ additions to obtain the 2^n coefficients $a_c(i)$ and $a_s(i)$ compared with the $2^n(2^n-1)$ additions necessary for the Walsh-Fourier transform.

THE WALSH - FOURIER TRANSFORM.

Consider a function F(θ) in some interval. Let this interval be divided into 2^n equally wide subintervals.

F(θ) is then a sample function having 2^n sample values.

The Walsh-Fourier transforms $a_c(i)$ and $a_s(i)$ of F(θ) are obtained by multiplying the samples with the values +1 or -1 of the first 2^n Walsh functions, summing the products and dividing the sum by the number- 2^n - of samples.

The same programme can be used for the forward and inverse transforms.

SEQUENCY AND FREQUENCY

The frequency of sinusoidal functions is usually defined as the number of cycles in a unit of time, but it would be just as appropriate to use half the number of zero crossings.

This latter interpretation is applied to Walsh functions, with the rate of oscillation being called sequency.

REPRESENTATIONS OF WALSH FUNCTIONS.

In the discussion above we introduced the Walsh functions using the notation wal(j,θ), or alternatively, wal(i,θ).

The first argument in this representation stands for sequency. Thus, in presenting the functions in the order wal(0,θ),wal(1,θ),....
we have given the sequency-ordered Walsh functions.

Chien(1975) gives an alternative representation of the Walsh functions which can be used to derive the explicit representation of a Walsh function of any order.

Walsh functions can also be represented as sums of Rademacher functions(Ohta, 1971,1976, and Harmuth, 1977, pp.44-6).

AS Chien(1975, p.170) has pointed out, it is a good sign that there are several different ways of defining and representing the Walsh functions since differing definitions may be most suitably used for particular applications.

Ahmed et al(1973) gives a summary of notation and representation for the Walsh functions.

HADAMARD MATRICES.

M is called an orthogonal matrix if the inverse M^{-1} and the transpose M^{t} are equal except for a factor.

An Hadamard matrix H is an orthogonal matrix with elements +1 and -1 only. (i.e. the pairwise correlation of any two rows equals zero).
There exists one Hadamard matrix each for ranks 1 and 2;

$$H_1 = \begin{bmatrix} +1 \end{bmatrix} \qquad H_2 = \begin{bmatrix} +1 & +1 \\ +1 & -1 \end{bmatrix}$$

Two non-trivially different matrices exist with a rank of 4;

$$H_{41} = \begin{bmatrix} +1 & +1 & +1 & +1 \\ -1 & -1 & +1 & +1 \\ -1 & -1 & +1 & +1 \\ +1 & -1 & +1 & -1 \end{bmatrix} \qquad H_{42} = \begin{bmatrix} +1 & +1 & +1 & -1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & +1 & +1 \\ -1 & +1 & +1 & +1 \end{bmatrix}$$

Matrices of higher rank may be obtained by the Kronecker products where the Kronecker product SXK of two matrices is obtained by multiplying S with each element k_{ij} of K and substituting the multiplied matrices $k_{ij}S$ for the elements k_{ij} of K.

USING KRONECKER PRODUCTS.

$$H_2 \times H_2 = \begin{bmatrix} H_2 & H_2 \\ H_2 & -H_2 \end{bmatrix} = \begin{bmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{bmatrix}$$

The resulting matrix is equal to H_{41} except for a different ordering of rows.

The rank of an Hadamard matrix with rank higher than two must be an integer multiple of four.

At least one Hadamard matrix is known for all possible ranks up to 200. Certain Hadamard matrices of rank 2^n are related to Walsh functions.

The Hadamard matrix of order 8 is given below (with the sequency of each row labelled).

$$H_8 = \begin{array}{cc} \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix} & \begin{array}{c} 0 \\ 7 \\ 3 \\ 4 \\ 1 \\ 6 \\ 2 \\ 5 \end{array} \end{array}$$

Each row represents the sampled values of wal(j,θ), where j is the sequency of the row.

The matrix can be sequency ordered by the appropriate row interchanges.

THE WALSH-HADAMARD TRANSFORM.

To obtain the Walsh-Hadamard transform of an $N \times N$ image $x(i,j)$, it is necessary to pre- and post-multiply by an N-th order symmetric Hadamard matrix.

The transformed image $y(k,l)$ is given by

$$Y = (1/N)\, HYH^t \quad \text{(matrix notation)}.$$

The inverse transform is

$$X = (1/N)\, HYH^t$$

Because of its binary nature, the Walsh-Hadamard transform can be efficiently implemented optically. The image is cross-correlated with a Walsh mask by placing the image transparency and the mask betwiin a light source and a detector.

TWO-DIMENSIONAL WALSH MASKS.

Figure 3 (after Harmuth, 1977, p.56) show the Walsh masks as they could be used to optically implement the Walsh-Hadamard transform. (Note however that the rows are ordered according to their sequency).

Instead of using the masks in image analysis, though, we are using them as a set of 64 stimuli in research which may perhaps suggest new possibilities in the fields of similarity, classification, and visual form perception in general.

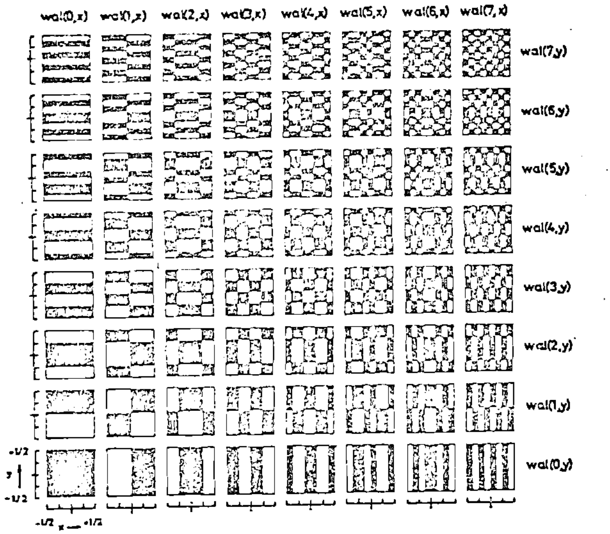


Figure 3. The Walsh masks in Cartesian co-ordinates.

SCHEMATIC REPRESENTATIONS OF THE WALSH STIMULI

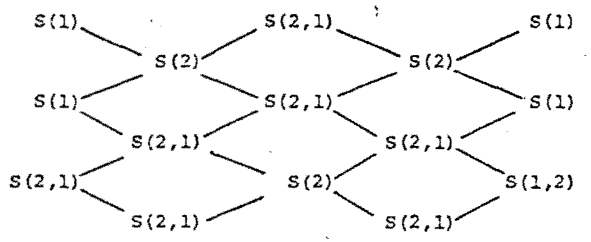
The 64 stimuli shown in figure three may be referred to by two numbers, the row and the column that they are in. Thus the stimulus in the top right hand corner may be described as Walsh 8,8 or Wal(8,8) and so on.

Alternatively the Walsh stimuli may be referred to using a notation which makes their structure explicit. Each of the 64 stimuli is made up of a number of patches or blobs, all of which are either square or rectangular in shape. For instance, Wal(8,8) is made up of 32 white squares and 32 black squares arranged as on a chess board.

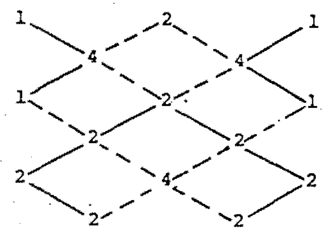
The Walsh stimuli may be constructed using various combinations of three sizes of square (1x1, 2x2, 4x4, 8x8) and six sizes of rectangle (2x1, 4x1, 8x1, 8x2, 8x4). For notational convenience, the shapes used may be labelled as follows:

Squares	Rectangles
1x1 = S(1)	2x1 = S(2,1)
2x2 = S(2)	4x1 = S(4,1)
4x4 = S(4)	4x2 = S(4,2)
8x8 = S(8)	8x1 = S(8,1)
	8x2 = S(8,2)
	8x4 = S(8,4)

The Walsh stimuli may now be represented as a set of shapes (using the above code) with links between adjacent shapes. Only the black shapes will be considered here, so that each shape is diagonally connected to the preceding shape. Using the present system, Wal(5,6) would be represented as follows:



A more convenient way of representating the Walsh stimuli is to use a graph theoretic approach with the node representing the area of the shape and the edge indicating whether the two nodes it connects have the same shape. Consider the first two shapes of Wal(5,6) reading from the top left hand corner. The size of the first shape may be denoted as N_S(left node) and will be equal to one. N_S(right node) the second shape will be four. Both these shapes are square, so a continuous edge may be used to indicate that the two nodes have the same shape. Thus in this graph-theoretic type of notation Wal(5,6) will be represented as:



This notation is particularly convenient in describing some of the features derived in Chapter two of this thesis.

Figure 4 (after Harmuth, 1977, p.61) shows the same 64 Walsh masks set out in polar, rather than cartesian, co-ordinates. Here we see another one of many different ways of representing the single underlying logical structure of the Walsh functions.

For the sake of uniformity of area, the scale divisions for the radius in figure 4 are located at the square roots of $\frac{1}{8}, \frac{3}{8}, \ldots, 1$.

$r^{\frac{1}{2}}$ is used instead of r in polar co-ordinates to ensure that the areas of the 64 underlying binary variables are mutually equivalent. (This is necessary because area increases proportionally with r^2.

TWELVE TEXTURAL FEATURES

Haralick et al (1973) considered the problem of deriving textural features for the purpose of image classification. Textural analysis is particularly important with the Walsh masks, since they have neither spectral nor contextural features, so that every feature that Walsh masks possess must in fact be closely related to texture. The fundamental unit of textural analysis is the spatial dependency matrix. Haralick et al devised some 14 textural features which could be derived from a given spatial dependency matrix. In similar vein, we calculated 5 features

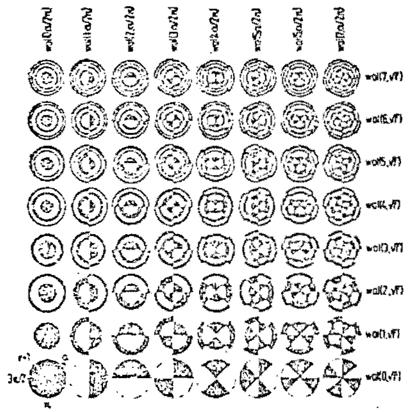


Figure 4. The Walsh masks in Polar co-ordinates.

on each of 12 spatial dependency matrices for each of the 63 non-homogeneous Walsh masks. The method for deriving the spatial dependency matrices is described in Haralick et al (1973) while a BASIC program to derive the 12 textural features is listed at the end of this appendix. Each Spatial Dependency matrix can be schematically represented in the following form:

Table B.1

		neighbour	
		level 1	level 2
reference square	level 1	P1	P2
	level 2	P3	P4

Table B.1: Schematic representation of a 2 gray level spatial dependency matrix where P1 is the proportion of times that both the reference square and its neighbour was black, while P4 represents the proportion of times when both squares were white. P2 refers to the proportion of black reference squares with white neighbours, while P3 refers to the proportion of white reference squares with black neighbours. The following featural statistics were calculated for each spatial dependency matrix:

1. $(P1)^2 + (P2)^2 + (P3)^2 + (P4)^2$.
2. $P2 + P3$
3. $[(P4*P1)-(P2*P3))/((P1+P2)\cdot(P1+P3)\cdot(P3+P4)\cdot(P2+P4)\}^{.5}$
4. $(P1-M)^2 + (P2-M)^2 + (P3-M)^2 + (P4-M)^2$
   where $M = (P1 + P2 + P3 + P4)/4$

5. $-[P1 \log(P1) + P2 \log(P2) + P3 \log(P3) + P4 \log(P4)]$

Thus feature 1 is the angular second moment, feature 2 is the sum of the off-diagonal entries, feature 3 is the negative of the sample value of ϕ in a four fold contingency table (Hays, 1973, p.743). Feature 4 is the sum of squares and feature 5 is a measure of entropy.

It turns out that features one, four and five are essential equivalent, as are features two and three.

We mentioned above that we calculated 12 spatial dependency matrices. These correspond to the Cartesian product of the set of angles $\{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ and the three distances $\{1,2,3\}$. Where d=1 implies nearest neighbour, d=2 implies next but one, and d=3 implies next but two.

The two features which we are considering are both functions of distance and angle. To produce rotationally invariant features, we have followed the suggestion of Haralick et al (1973) and taken the mean and range of each feature across the four angles $(0^\circ, 45^\circ, 90^\circ, 135^\circ)$ used.

We thus have a total of 12 textural features. These consist of:

a. The mean (across angles) phi-coefficient for each of the three dictances.

b. The range of the phi-coefficient over each of the three distances.

c. The mean entropy for each of the three distances.

d. The range of entropies for each of the three distances.

The values of these 12 textural features for 35 of the Walsh stimuli are given in Chapter Two (Table 2.8).

REFERENCES

Ahmed, N., Schreiber, H., and Lopresti, P.V., On notation and definitions of terms related to a class of complete orthogonal functions. IEEE Trans. Electromagn. Compat. EMC-15, 75-80(1973).

Chien, Ta-Mu. On representation of Walsh functions. IEEE Trans. electromagn. Compat. EMC-17 , 170-6(1975).

Haralick, R.M., Shanmugan, K. and Dinstein, I. Textural features for image classification. IEEE Trans. Syst., Man., Cyber., 1973, SMC-3, 610-621.

Harmuth, H.H. Sequency theory: foundations and applications. N.Y. Academic Press (1977).

Hays, W.L. Statistics for the Social Sciences. New York: Holt, Rinehart and Winston, (1973).

Ohta, T. The system of $r_n(\tau,t)$ functions - completion of the Rademacher system and periodic waveform synthesis. Electron. Commun. Japan 54, 33-40,(1971).

Ohta, T. Expansion of Walsh functions in terms of shifted Rademacher functions. IEEE Trans. Electromagn. Compat. EMC-18, (1976).

Welch, L.R. Computation of finite Fourier series. Space programs summary, Jet Propulsion Lab, Pasadena, California, 4, No. 37-39, 295-297. (1966).

Walsh, J.L. A closed set of normal orthogonal functions. Am. J. Math., 45, 5-24. (1923).

```
WALTEX  29-JAN-80  BASIC/CAPS  V01-01

2 REM     CALCULATE THE TEXTURAL FEATURES FOR THE WALSH MASKS
4 DIM W0(8,8),W(8,8),W8(8,8),W9(8,8)
7 DIM E0(64,2)
10 DIM V4(64,13),V3(64,13)
41 FOR T1=1 TO 64 \ FOR T2=1 TO 2
42 READ E0(T1,T2) \ NEXT T2 \ NEXT T1
45 N=8 \ M=8
60 Q$="****************************************************"
65 Q6$="======================================================="
2000 DATA 1,1,1,2,1,3,1,4,1,5,1,6,1,7,1,8
2001 DATA 2,1,2,2,2,3,2,4,2,5,2,6,2,7,2,8
2002 DATA 3,1,3,2,3,3,3,4,3,5,3,6,3,7,3,8
2003 DATA 4,1,4,2,4,3,4,4,4,5,4,6,4,7,4,8
2004 DATA 5,1,5,2,5,3,5,4,5,5,5,6,5,7,5,8
2005 DATA 6,1,6,2,6,3,6,4,6,5,6,6,6,7,6,8
2006 DATA 7,1,7,2,7,3,7,4,7,5,7,6,7,7,7,8
2007 DATA 8,1,8,2,8,3,8,4,8,5,8,6,8,7,8,8
2050 REM READ WOGS
2500 REM.......INITIALISE.........
2502 REM..............................
2510 FOR J1=2 TO 64
2515 FOR I1=1 TO 13
2520 V3(J1,I1)=0
2530 V4(J1,I1)=0
2540 NEXT I1
2550 NEXT J1
3000 FOR I=1 TO 8 \ FOR J=1 TO 8
3002 READ W0(I,J) \ NEXT J \ NEXT I
3020 FOR J=1 TO 8 \ FOR I=1 TO 8
3022 READ W(I,J) \ NEXT I \ NEXT J
3070 FOR D=1 TO 3
3072 PRINT "DISTANCE=  ";D \ PRINT  \ PRINT
3075 FOR J1=2 TO 64
3090 REM CELL STRUCTURE = WOGS MULTIPLIED
3100 FOR I=1 TO 8 \ FOR J=1 TO 8
3104 W8(I,J)=W0(E0(J1,1),I)*W(E0(J1,2),J)
3106 NEXT J
3108 NEXT I
3110 REM     CALCULATE THE FOUR SPATIAL DEPENDENCE
3115 REM..........MATRICES FOR A GIVEN D
3120 FOR Z8=1 TO 4
3122 P1(Z8)=0
3124 P2(Z8)=0
3126 P3(Z8)=0
3128 P4(Z8)=0
3130 ON Z8 GO TO 3140,3150,3160,3170
3140 M0=M-D
3141 N0=N
3142 I0=0
3143 J0=D
3144 S=1
3145 B0=1
3146 GO TO 3200
3150 M0=M-D
3151 N0=D+1
3152 I0=-D
3153 J0=D
3154 S=-1
3155 B0=8
3156 GO TO 3200
3160 M0=M
3161 N0=N-D
3162 I0=D
3163 J0=0
3164 S=1
3165 B0=1
3166 GO TO 3200
3170 M0=M-D
3171 N0=N-D
3172 I0=D
3173 J0=D
3174 S=1
3175 B0=1
3176 REM
3200 FOR I=B0 TO N0 STEP S
3205 IF I+I0<1 THEN 3275
3208 IF I+I0>8 THEN 3275
3210 FOR J=1 TO M0
3215 IF J+J0<1 THEN 3270
3218 IF J+J0>8 THEN 3270
3220 IF W8(I,J)<>W8(I+I0,J+J0) THEN 3260
3230 IF W8(I,J)=1 THEN P4(Z8)=P4(Z8)+1
3240 IF W8(I,J)=-1 THEN P1(Z8)=P1(Z8)+1
3250 GO TO 3270
3260 IF W8(I,J)=-1 THEN P2(Z8)=P2(Z8)+1
3265 IF W8(I,J)=1 THEN P3(Z8)=P3(Z8)+1
3270 NEXT J
3275 NEXT I
3280 REM.................NORMALISE...........
3290 R(Z8)=P1(Z8)+P2(Z8)+P3(Z8)+P4(Z8)
3291 IF R(Z8)=0 THEN 3299
3292 P1(Z8)=P1(Z8)/R(Z8)
3294 P2(Z8)=P2(Z8)/R(Z8)
3296 P3(Z8)=P3(Z8)/R(Z8)
3298 P4(Z8)=P4(Z8)/R(Z8)
3299 NEXT Z8
3300 GO TO 3800
3500 DATA 1,1,1,1,1,1,1,1
3501 DATA 1,1,1,1,-1,-1,-1,-1
3502 DATA 1,1,-1,-1,-1,-1,1,1
3503 DATA 1,1,-1,-1,1,1,-1,-1
3504 DATA 1,-1,-1,1,1,-1,-1,1
3505 DATA 1,-1,-1,1,-1,1,1,-1
3506 DATA 1,-1,1,-1,-1,1,-1,1
3507 DATA 1,-1,1,-1,1,-1,1,-1
3600 DATA 1,1,1,1,1,1,1,1
3601 DATA 1,1,1,1,-1,-1,-1,-1
3602 DATA 1,1,-1,-1,-1,-1,1,1
3603 DATA 1,1,-1,-1,1,1,-1,-1
3604 DATA 1,-1,-1,1,1,-1,-1,1
3605 DATA 1,-1,-1,1,-1,1,1,-1
3606 DATA 1,-1,1,-1,-1,1,-1,1
3607 DATA 1,-1,1,-1,1,-1,1,-1
3800 IF R(1)>0 THEN 4000
3805 IF R(2)>0 THEN 4000
3810 IF R(3)>0 THEN 4000
3820 IF R(4)>0 THEN 4000
3850 GO TO 5900
4000 REM      CALCULATE AND PRINT THE FEATURES FOR EACH ANGLE
4001 FOR Z8=1 TO 4
4002 GO TO 4025
4004 PRINT TAB(20);P1(Z8);TAB(32);P2(Z8)
4006 PRINT TAB(20);P3(Z8);TAB(32);P4(Z8) \ PRINT Q6$
4025 M1(Z8)=(P1(Z8)+P2(Z8))/2
4030 M2(Z8)=(P3(Z8)+P4(Z8))/2
4031 Q1(Z8)=P1(Z8)+P2(Z8)
```

```
4032 Q2(Z8)=P3(Z8)+P4(Z8)
4033 Q3(Z8)=P1(Z8)+P3(Z8)
4034 Q4(Z8)=P2(Z8)+P4(Z8)
4045 M3(Z8)=(P1(Z8)+P2(Z8)+P3(Z8)+P4(Z8))/4
4050 T3(Z8)=P2(Z8)+P3(Z8)
4055 D1(Z8)=P1(Z8)+P4(Z8)
4060 D2(Z8)=P2(Z8)+P3(Z8)
4065 F1(Z8)=P1(Z8)^2+P2(Z8)^2+P3(Z8)^2+P4(Z8)^2
4100 F2(Z8)=P2(Z8)+P3(Z8)
4107 F3(Z8)=((P1(Z8)+P2(Z8))*(P3(Z8)+P4(Z8)))
4110 F3(Z8)=F3(Z8)*(P1(Z8)+P3(Z8))
4112 F3(Z8)=F3(Z8)*(P2(Z8)+P4(Z8))
4115 IF F3(Z8)=0 THEN 4130
4120 F3(Z8)=(P1(Z8)*P4(Z8)-P2(Z8)*P3(Z8))/F3(Z8)^.5
4130 F4(Z8)=(P1(Z8)-M3(Z8))^2+(P2(Z8)-M3(Z8))^2
4140 F4(Z8)=F4(Z8)+(P3(Z8)-M3(Z8))^2+(P4(Z8)-M3(Z8))^2
4210 M5=P1(Z8)*(LOG(P1(Z8)+1.00000E-03)/LOG(2))
4215 M6=P2(Z8)*(LOG(P2(Z8)+1.00000E-03)/LOG(2))
4220 M7=P3(Z8)*(LOG(P3(Z8)+1.00000E-03)/LOG(2))
4222 M8=P4(Z8)*(LOG(P4(Z8)+1.00000E-03)/LOG(2))
4225 F9(Z8)=-(M5+M6+M7+M8)
4490 NEXT Z8
5000 REM  CALCULATE AND STORE THE 10 ROTATIONALLY INVARIANT FEATURES
5002 FOR I1=1 TO 5
5003 V3(J1,I1)=0
5004 V1(I1)=1.00000E+06
5005 V4(J1,I1)=0
5006 V2(I1)=-1.00000E+06
5008 NEXT I1
5010 FOR Z8=1 TO 4
5020 V4(J1,1)=V4(J1,1)+F1(Z8)
5022 V4(J1,2)=V4(J1,2)+F2(Z8)
5024 V4(J1,3)=V4(J1,3)+F3(Z8)
5026 V4(J1,4)=V4(J1,4)+F4(Z8)
5028 V4(J1,5)=V4(J1,5)+F9(Z8)
5060 IF F1(Z8)<=V1(1) THEN V1(1)=F1(Z8)
5062 IF F1(Z8)>=V2(1) THEN V2(1)=F1(Z8)
5064 IF F2(Z8)<=V1(2) THEN V1(2)=F2(Z8)
5066 IF F2(Z8)>=V2(2) THEN V2(2)=F2(Z8)
5068 IF F3(Z8)<=V1(3) THEN V1(3)=F3(Z8)
5070 IF F3(Z8)>=V2(3) THEN V2(3)=F3(Z8)
5072 IF F4(Z8)<=V1(4) THEN V1(4)=F4(Z8)
5074 IF F4(Z8)>=V2(4) THEN V2(4)=F4(Z8)
5076 IF F9(Z8)<=V1(5) THEN V1(5)=F9(Z8)
5078 IF F9(Z8)>=V2(5) THEN V2(5)=F9(Z8)
5150 NEXT Z8
5200 REM.......
5205 REM   CALCULATE THE MEANS AND RANGES
5210 REM.........................................
5220 FOR I1=1 TO 5
5222 V4(J1,I1)=V4(J1,I1)/4
5225 V3(J1,I1)=V2(I1)-V1(I1)
5300 NEXT I1
6100 NEXT J1
7000 REM   OUTPUT THE RESULTS IN CONDENSED FORMAT
7050 FOR J1=2 TO 64
7060 PRINT "MASK";J1;
7100 FOR I1=1 TO 5        MEANS
7110 PRINT TAB(I1*12);V4(J1,I1);
7120 NEXT I1
7125 PRINT
7130 FOR I1=1 TO 5        RANGES
7140 PRINT TAB(I1*12);V3(J1,I1);
7150 NEXT I1
7160 PRINT
7200 NEXT J1
7300 NEXT D
8000 END
```

— I1 = 3, 5 generated the features used in this thesis.

F12 → D=1, I1=3, MEAN
F13 → D=1, I1=5 MEAN
F14 → D=1, I1=3, RANGE
F15 → D=1, I1=5, RANGE
F16 → D=2, I1=3 MEAN
⋮
F25 → D=3, I1=5, RANGE.

APPENDIX C

This appendix will outline a technique called conceptual ranking which may be used as an heuristic method for unidimensional scaling.

Cliff and Young (1968) concluded from a set of three studies that there was "a high degreee of consistency between single-stimulus judgements and spaces derived from MSA (multidimensional scaling analysis) of similarities judgements". However, Green and Cormone (1970) did not obtain good agreement between single and multidimensional judgements and Zinnes and Wolff (1977) found that multidimensional stimuli were more indiscriminable than expected on the basis of unidimensional judgements of these stimuli. The relationship between unidimensional judgements and multidimensional scaling is thus far from clear.

Two methods of unidimensional scaling are rating and rankings. Cook and Smith (1974) argued that rankings were preferable to ratings because ranking eradicates differences in level and spread which in an analysis of ratings are simply nuisance factors. Experiment E1 (reported in chapter two of this thesis) was a graphic example of the marked differences in level and spread of ratings which can occur between participants.

One approach to the development of novel ranking techniques in psychological data collection is to consider the sorting algorithms used by computer scientists (e.g. Flores, 1969). Another approach is to use quick but crude methods, such as the one to be described below, and then test the usefulness of these methods empirically.

No attempt will be made to give a mathematical development of the technique here, although a general formalisation of ranking techniques based on Young tableaux (Rutherford, 1948) should eventually be possible. A Young tableau of shape $(n_1, n_2, n_3, \ldots n_m)$, where $n_1 \geqslant n_2 \geqslant \ldots \geqslant n_m \geqslant 0$, is an arrangement of $n_1+n_2+\ldots\ldots+n_m$ distinct integers in an array of left-justified rows, with M_i elements in row_i, such that the entries of each row are in increasing order from left to right, and the entries of each column are increasing from top to bottom. While a Young tableau does not define a unique ordering of a set of objects, it does define an approximate ordering which can be derived from the expected values of the integers in each cell of the tableau. In the following development of the conceptual ranking technique, the psychological rather than mathematical properties of the method will be stressed, although it is noted here that a completed two-way conceptual grid is in fact an upside-down Young tableau.

Efficient Procedures for Conceptual Ranking.

We shall consider here methods which are appropriate for stimuli which can be physically manipulated and sorted.

The One-way Conceptual Rank.

The first method is to rank order stimuli from 1 to 64. This method has been largely ignored in the past, and with good reason.

Two of the problems which arise are:

(i) It is an extremely messy task and the participant typically finds it difficult to consistently use a single criterion when dealing with 64 stimuli at once.

(ii) 64 ordinal ranks are obtained from the method, but it is doubtful whether the method can distinguish more than about seven levels of a concept.

The Two-way Conceptual Rank

The second method is as follows:

The 64 objects are randomly placed into the form of an 8 by 8 grid on the top of a large table. The participant then carries out the following procedure.

Step 1.

Rank the first column (working left to right, say) in ascending order (working upwards) according to the value of the concept.

Then, do the same with each of the seven remaining columns, working from left to right across the columns.

Step 2.

Rank the first (bottom, say) row in ascending order (working to the right) according to the value of the concept. Do the same within each of the seven remaining rows, working up the grid row by row.

Step 3.

If none of the stimuli have been moved since you were last at STEP 3, then Got to STEP 4. Otherwise, return to STEP 1.

Step 4.

Are your satisfied with the configuration?

The result of this procedure is an 8 by 8 grid of the objects which is ranked simultaneously over both rows and columns by the value of the concept.

Three-way Conceptual Ranking

The third method is as follows:

The participant sorts the 64 stimuli into four piles, with each pile representing a level of the concept. He or she then adjusts the piles so that each of the four levels contains 16 stimuli.

A 4 by 4 (2-way) conceptual ranking is then carried out on each of the four groups of 16 stimuli.

The Generalized Conceptual Rank

Higher N-way conceptual rankings can be constructed but their utility in psychological research is doubtful. The generalized conceptual rank is the three-way rectangular (N x M1 x M2) conceptual ranking where $N \times M1 \times M2 \leq$ the number of stimuli. In this case, the data are first sorted into N levels, and then 2-way conceptual ranks are then constructed on an M1 and M2 rectangular grid for each level.

The number of levels of complexity generated in such a procedure varies with the metric assumed to hold across the conceptual rank. For a city-block metric there would be N (M1 + M2 - 1) levels. 2-way and 1-way conceptual ranks are special cases of the three-way procedure and can be described as 1 x M1 x M2 - and N x 1 x 1 rankings, respectively.

How Conceptual Ranking can increase Information Transmission

The power of the conceptual ranking procedure is that it forces the participant to judge each stimulus against the other stimuli. Thus it is a relative judgement procedure but is far less tedious than a procedure such as pairwise comparisons which requires ${}^{64}C_2$ trials (in our example) for a complete design. Furthermore, the total ranking task is segmented into a series of successive operations, so that the participant only had to judge the relative rank of eight stimuli at any one time.

A similar method, which could be used with stimuli (auditory stimuli for instance) which cannot be physically manipulated by the participant, is to rank the stimuli into a small number of levels and then successively rank within each of the levels until a complete rank order is derived. However, this method fails to utilise the main advantage of physically manipulable stimuli which is that they can be rearranged within a two-dimensional configuration. The method of successive ranking forms nested hierarchies, and, as a result there is no way of moving a stimuli again once it has been put into a particular level.

At this stage conceptual ranking appears to be an efficient way of transmitting information between the participant and the experimenter since the physical rearrangement of the stimuli is the experimental response.

Aside from the data analysis problems which we shall consider below, conceptual ranking should reduce at least three sources of noise in the experiment.

DERIVING UNIDIMENSIONAL SCALES FROM A CONCEPTUAL RANKING

In the absence of an appropriate statistical theory Monte Carlo simulation techniques were used to estimate the ranks of objects placed in a conceptual grid.

The following set of tables (Tables C.1 to C.10) give the mean rank and the standard deviation of the rank obtained over 100 simulation trials (using the conceptual ranking procedure with initially random configurations) for each cell in the resulting conceptual grids. The results are for 3x3, 4x4, 5x5, 6x6, and 7x5 conceptual grids.

	column		
Row	1	2	3
1	1.0	2.2	4.4
2	3.2	4.8	6.8
3	6.0	7.8	9.0

Table C.1 Expected rankings in a 3x3 Conceptual grid

	column		
Row	1	2	3
1	.00	.36	1.11
2	.65	.68	.73
3	.99	.45	.00

Table C.2 Standard deviations of the rankings for each position in the grid (obtained in 100 simulation runs)

Row	column 1	2	3	4
1	1.0	2.2	4.3	7.8
2	3.2	5.3	7.6	11.1
3	5.9	8.6	10.7	13.3
4	10.8	13.3	14.9	16.0

Table C.3 Expected rankings in a 4x4 conceptual grid

Row	column 1	2	3	4
1	.00	.45	1.28	1.98
2	.79	.88	1.09	1.39
3	1.46	1.13	1.06	.87
4	1.76	.92	.32	.00

Table C.4 Standard deviations of the rankings for each position in the grid (obtained in 100 simulation runs using a 4x4 grid

Row	column 1	2	3	4	5
1	1.0	2.3	4.4	7.3	11.9
2	3.2	5.7	8.2	11.8	17.0
3	5.9	9.5	11.9	14.6	20.1
4	10.3	13.9	16.5	18.8	22.2
5	17.4	20.3	22.2	23.8	25.0

Table C.5 Expected values in a 5x5 conceptual grid

Row	column 1	2	3	4	5
1	.00	.69	1.27	2.26	3.54
2	.92	1.11	1.29	1.67	2.20
3	1.78	1.42	1.40	1.53	1.74
4	2.81	1.82	1.29	1.39	1.15
5	2.35	1.61	1.06	.52	.00

Table C.6 Standard deviations of the rankings for each position in the grid (obtained in 100 simulation runs using a 5x5 grid)

Row	column 1	2	3	4	5	6
1	1.0	2.4	4.3	7.1	11.3	18.5
2	3.2	6.1	8.7	12.3	18.1	25.5
3	5.8	10.2	12.9	16.0	21.6	28.6
4	9.5	14.6	17.7	20.6	24.5	31.0
5	14.7	19.8	23.4	26.3	28.7	33.2
6	24.5	28.6	31.4	33.2	34.7	36.0

Table C.7 Expected values in a 6x6 conceptual grid

Row	column 1	2	3	4	5	6
1	.00	.60	1.38	2.18	3.60	5.40
2	1.07	1.33	1.57	2.17	2.91	2.93
3	1.85	1.77	1.66	1.78	2.26	2.34
4	2.63	2.16	1.93	1.86	2.02	1.75
5	3.75	2.57	2.06	1.62	1.85	1.40
6	3.92	2.42	1.56	1.07	.74	.00

Table C.8 Standard deviations of the rankings for each position in the grid (obtained in 100 simulation runs using a 6x6 grid

Row	column 1	2	3	4	5	6	7
1	0.000	0.168	1.264	1.957	2.671	3.623	3.496
2	0.987	1.373	1.549	1.862	2.256	2.499	1.944
3	1.785	1.982	1.763	1.765	1.734	1.919	1.388
4	2.756	2.886	2.511	1.943	1.883	1.692	0.740
5	4.743	3.183	1.988	1.495	1.212	0.989	0.000

Table C.9 Standard deviation in ranks for each position of a 7x5 conceptual grid

Row	column 1	2	3	4	5	6	7
1	1.00	2.41	4.39	7.03	10.63	16.07	25.09
2	3.08	5.88	8.96	12.05	16.10	21.42	29.14
3	5.75	10.48	13.48	16.19	19.55	24.67	31.71
4	9.32	15.35	18.88	21.73	24.15	27.72	33.65
5	16.61	23.01	26.70	29.06	31.01	32.73	35.00

Table C.10 Expected rank for each position in a 7x5 conceptual grid

1. Numerical response bias is removed by forcing participants to use the same physical ordination procedure.
2. The procedure requires a single session of forty minutes which reduces the fatigue and boredom which inevitably result from long multi-trial experiments.
3. Conceptual ranking does not require speeded responding, since the participant is in complete control of the task once the initial configuration has been presented and the procedure explained. Furthermore, the experiment ends only when the participant is happy that his conceptual ranking fulfils the condition of monotonic increase in the concept over <u>all</u> rows and columns. If the participant is unhappy with his final configuration then the experiment is abandoned (this has not happened with the 12 participants used so far).

The method of conceptual ranking certainly deserves more attention in the field of visual form perception. It appears that it can be used to rapidly and reliably scale a set of from twenty to a hundred visual stimuli in terms of some dependent variable. It is worthwhile contrasting this with the approaches taken in a recent paper on the complexity of visual patterns Chipman, 1977). In this important paper three main data collection methods were used:-

1). Magnitude Estimation.

The geometric mean of judgements was used as the basic datum, making individual scaling impossible. In addition, it was found necessary to replicate the task so as to prevent non-stationarity in the use of numbers over time.

2). Pairwise comparisons.

One experiment used only eight stimuli, the other was an incomplete and unbalanced design for 37 stimuli which still required 253 judgements.

3). One-Way Ranking.

Only 17 stimuli were used, but this is still a messy task. Furthermore the results were converted to implied pair-comparisons before analysis.

Chipman's paper illustrates the current methodologies in visual form perception. The generalised experimental design considered above should be able to deal with some of the unresolved difficulties in current methodology. These include the imprecise specification of the 'to-be-judged' attribute, the general absence of individual scaling, tedious experimental tasks with associated non-stationarity problems (Gregson, 1974a,b), numerical response biasses, and the use of the general linear model or the minkowski metric in multidimensional scaling without attempting to design alternative models which might be more relevant to the situation.

References

Cliff, N. and Young, F.W. (1968) On the relation between unidimensional judgements and multidimensional scaling. Organizational Behavior and Human Performance, 3, 269-285.

Cook, M. and Smith, J.M.C. (1974) Group ranking techniques in the study of accuracy of interpersonal perception. British Journal of Psychology, 65, 427-435.

Flores, I. (1969) Computer Sorting. Englewood Cliffs: Prentice-Hall.

Green, P.E. and Carmone, F.J. (1970) Multidimensional Scaling and Related Techniques in Marketing Analysis. Boston: Allyn and Bacon.

Gregson, R.A.M. (1974a) Similarity judgements modified by feedback. Acta Psychologica, 38, 117-129.

Gregson, R.A.M. (1974b) Sequences of similarity judgements as time series. Acta Psychologica, 38, 429-445.

Rutherford, D.E. (1948) Substitutional Analysis. New York: Hafner.

Zinnes, J.L. and Wolff, R.P. (1977) Single and multidimensional same-different judgements. Journal of Mathematical Psychology, 16, 30-50.

APPENDIX D

Chapters one and eight of this thesis briefly considered the problem of what form stored representations take. This appendix will lock at some methods which can be used to detect particular types of stored representation.

One method which has been used is that of looking for signs of clustering and organisation in free recall data (Shuell, 1969). The rationale for this method is that the organisation of the stored representations will be reflected in the organisation of the recalled items.

A second method looks for signs of clustering and organisation in sorting data, which was the method used in chapter eight of this thesis.

A third method uses clustering and other organisation in similarities data as estimates of underlying cognitive structure (Shepard, 1974). It was suggested in chapter five of this thesis that this third type of method was more likely to be an estimate of representational structure than an estimate of the stored representations.

A common feature of the three methods outlined above is that the cognitive metatheory implies the use of cluster analysis in estimating underlying cognitive structure. Miller (1969) used complete-linkage and single-linkage hierarchical clustering in estimating structure from sorting data, whereas average-linkage hierarchical clustering was used in chapter eight of the present thesis. One of the supposed advantages of the complete-linkage methods is that they remain invariant to monotone transformations of the input data (Milligan, 1979). The method of average linkage (otherwise known as the group average, or UPGMA, method) was used, however, in this thesis as it is a compromise between the "space-dilating" effect of single-linkage clustering and the "space-contracting" effect of complete-linkage clustering (Everitt, 1974).

The clustering methods mentioned above were originally developed to deal with problems in numerical taxonomy (Sokal and Sneath, 1963).

Johnson's(1967) paper can be seen as the beginning of an upsurge in interest in clustering analysis from a psychological perspective. This has culminated in the recent development of the ADCLUS (Shepard and Arabie, 1979) and INDCLUS(Carroll and Arabie, 1979) procedures which allow the representation of similarities as combinations of discrete overlapping properties (with separate fitting of weights on each property, for each participant in the case of INDCLUS).

An alternative approach to the problem of representing objects which each have a number of properties (i.e. belong to a number of clusters) is to use fuzzy cluster analysis. As this technique is unfamiliar to most psychologists, the remainder of this appendix constitutes an introduction to fuzzy cluster analysis.

BRIEF INTRODUCTION TO FUZZY CLUSTER ANALYSIS

The inherent fuzziness of psychological concepts can now be quantified using the theory of fuzzy sets (first outlined by Zadeh, 1965).
Fuzzy sets are particularly useful in classification problems because the variety of substructures and inter-relationships existing in real data precludes the possibility of finding a single criterion capable of identifying 'optimal' partitionings for a particular data set.

The field of classification and clustering has been comprehensively reviewed in Duda and Hart(1973), and Tou and Gonzalez(1974), however it is only recently that fuzzy algorithms have become available.

The evolution of fuzzy sets as a theoretical basis for cluster analysis can be traced through the following sequence of papers;-

Bellman et al(1966), Wee(1967), Flake and Turner(1968), Gitman and Levine(1970), Ruspini(1969,1970), and Dunn(1974a,1974b).

Ruspini(1969) delineated the first objective function method for fuzzy clustering. His algorithms are given in Ruspini(1970,1972,1973).

The fundamental conceptual advance on both fuzzy clustering and fuzzy set theory is to allow each element to have grades of membership with respect to different sets rather than just having the dichotomy of absence or presence.

We shall discuss here the algorithm of Dunn and Bezdek(Bezdek, 1973, and Dunn, 1974a). In particular we shall draw on a number of recent papers by Bezdek (1974a,1974b,1975,1976,1977) for our exposition.

DEFINITION

Let V_{cn} denote the vector space of all real (c x n) matrices, and let u_{ik} be the ik-th element of $U \in V_{cn}$. Every disjoint c-partition of X can be uniquely represented by a matrix in the set:-

$$M_c = \left\{ U \in V_{cn} : u_{ik} \in \{0,1\}\ \forall i,k ; \sum_{i=1}^{c} u_{ik} = 1\ \forall k ; \sum_{k=1}^{n} u_{ik} > 0\ \forall i \right\}$$

Fuzzy clustering involves extending the range of each characteristic function (U_{ik}) from $\{0,1\}$ to $[0,1]$ so that each case now has a grade of membership in each cluster. It is important to note that the grade of membership is not a probability but a measure of how closely a profile agrees with the characteristic pattern associated with the i-th subclass.

Extending the range of the membership function means that one object profile can belong (with varying grades of membership) to a number of different clusters. This enables the fuzzy clustering to represent a variety of substructures and inter-relationships existing within a data set.

In any clustering technique convergence is determined by some global criterion function (i.e, by a goodness-of-fit type measure). Dunn and Bezdek have used the following;

$$[1] \qquad J_m(U, \underline{v}) = \sum_{k=1}^{n} \sum_{i=1}^{c} (u_{ik})^m \| \underline{x}_k - v_i \|^2 ; \quad 1 \le m < \infty$$

where u_{ik} is the membership function for the k^{th} case in the i-th cluster, $\underline{x}_k$ is the data vector for the k^{th} case

and $\underline{v}_i$ is the cluster centre for the i^{th} cluster.
$\| \|$ is any differentiable norm on R^s -in our own particular case we shall only consider the Euclidean norm.
Bezdek has called the resulting family of clustering algorithms for iterative optimisation of J_m -the fuzzy ISODATA algorithms.

Recent applications of fuzzy ISODATA for real data processing include pharmacology (Rossini et al, 1975), and medical diagnosis (Fordon and Fu, 1976).
Fordon and Fu, for example, state that fuzzy ISODATA appears to reduce the error rate in detection of false renovascular patients in hypertension studies by a factor of two over all previously tried methods.

THE ALGORITHM.

Given a set of N data vectors $\underline{x}_k$ (each consisting of D variables) which we shall denote X (= $\underline{x}_1, \ldots, \underline{x}_n$) fuzzy ISODATA generates fuzzy c-partitions of X by iteratively optimising the fuzzy least squared error functional given in 1 .

The following equations provide a loop for iterative minimesation of I .

$$(2a) \qquad \hat{u}_{ik} = \frac{1}{\sum_{j=1}^{c} \left[\frac{\|\underline{x}_k - \hat{\underline{v}}_i\|}{\|\underline{x}_k - \hat{\underline{v}}_j\|} \right]^{2/m-1}} \qquad 1 \le i \le c, \; 1 \le k \le n$$

$$(2b). \qquad \hat{\underline{v}}_i = \frac{\sum_{k=1}^{n} (\hat{u}_{ik})^m \underline{x}_k}{\sum_{k=1}^{n} (\hat{u}_{ik})^m} \, , \forall i .$$

Equations (2a) and (2b) are necessary if $(\hat{U}, \hat{\underline{V}})$ is to minimise the criterion, J_m, given in [1] .
This was shown by Grim et as (1977).

THE ITERATIVE PROCESS.

1/. Fix m, $1 \le m < \infty$(m is a weighting exponent. If m=1, then we have the usual (hard) k-means algorithm. As m gets larger than 1 the solution gets fuzzier).

2/. Choose any c x n matrix U_o. (This is the initial guess for the membership functions over the c clusters and the n data vectors).

3/. Calculate the weighted means ($\hat{\underline{V}}_i$) for each of the c clusters using equation (2b). These then become the current cluster centres.

4/. Update the membership functions using equation (2a). Replace U_o with U .

5/. Compute the maximum membership defect;
$$\max_{i,k} \{ | (\hat{u})_{ik} - (u_o)_{ik} | \}$$
Compare this 'defect' with a cutoff criterion e.
If $\| U_o - U \| \le e$, then stop.
Otherwise, put $U \rightarrow U_o$ and go to 3.

Equations (2a) and (2b) are the crux of the FUZZY ISODATA process. Dunn (1974a) gives the details for m=1 and the singular cases $\underline{x}_k = \underline{v}_k$ for some i and k. For m=1 the resulting algorithm is essentially the hard ISODATA process of Ball and Hall (1967).

Equation (2a) can be seen to describe a comparative proximity measure where grade of membership in a cluster is inversely proportional to the distance of that cluster relative to the weighted average cluster distance.

In equation (2b) the contribution of each profile to the cluster centre is directly proportional to its grade of membership in that cluster relative to those of all the other profiles.

It can thus be seen that this technique conforms to basic notions that any classifying process involves two fundamental procedures:-

1/. Comparative judgements between the objects to be classified.

2/. Comparative judgements between the categories to be used in classification.

Consequently, as well as being a useful analytic technique, the fuzzy ISODATA algorithm also embodies a powerful psychological theory on the classifying process itself.

CLUSTER VALIDITY

Unfortunately, global minima of objective functions such as J_m may suggest very poor interpretations of substructures in X (the n x d data matrix), as shown in Bazdek (1974b), Wishart (1969), and Ling (1971).

In contrast to conventional clustering, however, the fuzziness of U (the c x n matrix of membership functions) allows one to associate various measures of partition quality with U which are independent of the method used to produce these partitions.

This allows us to evaluate the validity of c - the number of subclasses hypothesised, in much the same way as indices of stress are used to determine the appropriate dimensionality in multidimensionality scaling.

The two measures of partition fuzziness that we are concerned with are:-

a) $F_i(U) = \mathrm{trace}(UU^T)/n$; Partition Coefficient
Bezdek (1974b)

This is generally used in the form $1-F_i(U)$.

b) $H_c(U) = -(\sum_{b=1}^{n}\sum_{i=1}^{c} U_{ik} \log_a U_{ik})/n$, with $a \varepsilon (1,)$.

This is generally used with natural logs and is called Entropy, Bezdek (1975).

Theoretical results appearing in Bezdek (1975) suggest that minimization of H_c of $1-F_i$ leads to an optimal choice for $U \varepsilon M_i$. The accompanying numerical experiments suggested that H_c is a somewhat sharper measure than F_i.

To allow for spurious fitting due to a large number of clusters, H_c is normalised by having

$$H_{c_{norm}}(U) = H_c(U)/(1 - C/N)$$

This is the measure that we have been using.

We have now covered the essential elements of this particular fuzzy clustering technique. We shall finish by discussing a couple of issues concerning the use of the technique.

Initial Guesses

To start the algorithm off, we need an initial guess of the U_{cn} matrix of membership functions. As with all hill-climbing procedures, fuzzy ISODATA is sensitive to initial guesses. Ideally, several different starting guesses should be used to guard against stagnation at a local minimum of the objective function.

In our work we have so far only used the following initial guess of Bezdek (personal communication) which we have found to work quite satisfactorily.

$$\text{let } U = \begin{array}{cc} \begin{matrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & 0 & \cdots & 1 \end{matrix} & \begin{matrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & 0 & 0 & \cdots & 0 \\ 1 & \cdot & & & \vdots \\ 1 & \cdot & & & \vdots \\ 1 & 0 & 0 & \cdots & 0 \end{matrix} \\ c\times c & c\times(n-c) \end{array}$$

let B be a matrix (cxn) with every element equal to $\frac{1-\sqrt{\frac{1}{2}}}{c}$.

Then U_o (the initial guess) $= \sqrt{\tfrac{1}{2}}.U + B$.

The initial guess given here is very similar to that of Bezdek (1976). The actual form of U_o is chosen to be between U (given above) and the disjoint c-partition of X, as measured by the value of F_c.

Sequential Fuzzy Clustering

We have seen that the fuzzy clustering process requires a number of parameters to be set before it can be used.

1. the number of clusters, c.
2. the weighting experiment, m.
3. the initial guess, U_o.
4. the appropriate norm. This is analogous to the choice of metric in multidimensional scaling. The Euclidean norm is a good compromise but there are other possibilities e.g. the Mahalanobis norm.
5. the maximum defect .
6. the maximum number of iterations.

Bezdek (1974a) shows how to conduct a systematic search for structure using fuzzy clustering. Unfortunately though, fuzzy clustering becomes prohibitively expensive (in computing time) when the number of clusters exceeds 4 and D, the number of variables (features) exceeds 30. For instance, when searching for 9 clusters using 27 cases and 32 features, we found that each iteration was taking about 6 hours on a PDP11/10. Therefore for many psychological applications I suggest the following procedure:-

a) use the following parameter settings; m = 1.25, = .01, guess as given here, maximum number of iterations = 25. Fuzzy cluster analysis can be used on a minicomputer (we have been using it on a PDP11/40), but as a rule of thumb I suggest that no single analysis should be left more than 24 hours - the extra information just isn't worth it.

These parameter settings seem a good compromise considering that the technique is new and no firm guidelines have been set down in the literature.

b) vary the number of clusters, starting with C = 2. Because the membership functions are constrained to sum to one for each case, the cluster solutions can be graphed in C -1 dimensions. Thus I think one should routinely plot the solutions for C = 2,3,4.

If C = 4 given the best fit, then, and only then, considering using larger Cs, always balance further exploration against the cost of increased computer time.

Conclusion

Fuzzy cluster analysis should be an important addition to the armoury of analytic weapons available to the psychologist. As well as being a reasonably unbiased method for exploring the structural possibilities within a set of data it is also a formalisation for the important new metatheory of fuzzy sets which has much to offer psychology in the future. Unfortunately, I do not expect it to be extensively used by psychologists in the near future because of their obvious preference for "cut and dried" techniques. The avoidance of Bayesian analysis on the one hand, and the over-use of factor analysis and t-tests on the other are indicative of an unwillingness to tolerate the notion of uncertainty in experimental data. Fuzzy clustering is an interesting new technique precisely because it makes this uncertainty explicit.

References

Ball, G.H. and Hall, D.J. (1967). A clustering technique for summarizing multivariate data. Behavioral Science, 12, 153-155.

Bellman, R.E., Kalaba, R., Zadeh, L. (1966) Abstraction and pattern classification. J. Math. Anal. Appl.; 13, 1-7.

Bezdek, J.C. (1973). Fuzzy maths in pattern classification. Thesis: Ph.d. Cornwall University, Ithaca, New York.

Bezdek, J.C. (1974a). Numerical taxonomy with fuzzy sets. J. Math. Biol., 1, 57-71.

Bezdek, J.C. (1974b) Cluster validity with fuzzy sets. Journal of Cybernetics, 3, 58-73.

Bezdek, J.C. (1975) Mathematical models for systematics and taxonomy. pp. 143-166 in Estabrook, G.(Ed.), Proceedings of the Eighth International Conference on Taxonomy. San Francisco: Freeman.

Bezdek, J.C. (1976) Feature selection for binary data: Medical diagnosis with fuzzy sets. Conf. Proceedings of National Computers Conference: 1976, June 7-10, NY, NY, AFIPS Press, Montrale, NJ.

Bezdek, J.C. (1977) Some current trends in fuzzy clustering. Proc. of the Annual North American Meeting of the Society for General Systems Research, published by SGSR, Washington, DC, February, 1977, 347-352.

Bezdek. J.C. and Fordon, W.A. Analysis of Hypertensive Patients by the use of the fuzzy ISOPATA algorithm. (unpublished manuscript).

Carroll, J.D. and Arabie, P. (1979). INDCLUS: A three-way approach to clustering. Paper presented to the Psychometric Society on June 9, 1979.

Duda, P. and Hart, P. (1973). Pattern Classification and Scene Analysis. NY: Wiley.

Dunn, J.C. (1974a) A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. Journal of Cybernetics, 3, 32-57.

Dunn, J.C. (1974b) Well-separated clusters and optimal fuzzy partitions. Journal of Cybernetics, 4, 95-104.

Everitt, B.S. (1974) Cluster Analysis. London: Heinemann.

Flake, R.H. and Turner, B.L. (1968) Numerical classification for taxonomic problems. J. Theoretic Biol., 20, 260-270.

Fordon, W. and Fu, K.S. (1976) Computer-aided differential diagnosis of hypertension. Technical report: Electrical Engineering, 76-77, Purdue University, Lafayette (August), 110-123.

Gitman, I. and Levine, M. (1970) An algorithm for detecting unimodal fuzzy sets and its application as a clustering technique. IEEE Trans on Computers C-19, 583-593.

Grim, C.E. Weinberger, M.H. et al (1977) Diagnosis of secondary forms of hypertension: A comprehensive protocol. J. American Medical Association, 237, 1331-1335.

Johnson, S.C. (1967) Hierarchical clustering schemes. Psychometrika, 32, 241-254.

Ling, R.F. (1971) Cluster analysis. Thesis: Ph.d. Yale University, New Haven.

Miller, G.A. A psychological method to investigate verbal concepts. Journal of Mathematical Psychology, 1969, 6, 209-220.

Milligan, G.W. (1979) Ultrametric hierarchical clustering algorithms. Psychometrika, 44, 343-346.

Rossini, L., Matorana, F. and Periti, P. (1975) Clustering cholinergic receptors by muscarine and muscarone analogs, in Proceedings of the 2nd International Meeting of Medical Advisors in the Pharmaceutical Industry (October) : Firenze, Italy.

Ruspini, E. (1969) A new approach to fuzzy clustering. Information and Control, 15, 22-32.

Ruspini, E. (1970) Numerical methods for fuzzy clustering. Information Science, 2, 319-350.

Ruspini, E. (1972) Optimization in sample descriptions: Data reduction and pattern recognition using fuzzy clustering. IEEE Trans. on Syst., Man., and Cybernetics SMC-2, 541.

Ruspini, E. (1973) New experimental results in fuzzy clustering. Information Science, 6, 273-284.

Shepard, R.N. (1974) Representation of structure in similarity data: Problems and prospects. Psychometrika, 39, 373-421.

Shepard, R.N. and Arabie, P. (1979) Additive clustering: Representation of similarities as combinations of discrete overlapping properties. Psychological Review, 86, 87-123.

Shuell, T.J. (1969) Clustering and organizaiton in free recall. Psychological Bulletin, 72, 353-374.

Sokal, R.R. and Sneath, P.H.A. (1963) Principals of Numerical Taxonomy. San Francisco: Freeman.

Tou, J. and Gonzalez, R. (1974) Pattern Recognition Principals. Reading: Addison-Wesley.

Wee, W. (1967) On generalizations of adaptive algorithms and applications of the fuzzy sets concept to pattern classification. Thesis: Ph.d. Purdue University, Lafayette, Indiana.

Wishart, D. (1969) Mode analysis: A generalization of nearest neighbour which reduces chaining effects in Numerical Taxonomy, Cole, A. (Ed.), Academic Press, NY, pp. 282-308.

Zadeh, L.A. (1965) Fuzzy sets. Information and Control, 8, 338-353.